A Bootstrap Procedure for Inference in Nonparametric Instrumental Variables

Andres Santos

Department of Economics

University of California, San Diego

e-mail: a2santos@ucsd.edu

May, 2008

Abstract

This paper proposes a consistent bootstrap procedure for the test statistic of Santos (2007). The derived bootstrap allows for inference in partially identified nonparametric instrumental variables models. It can be employed to test whether at least one element of the identified set satisfies a conjectured restriction. Possible applications include testing for shape restrictions such as economies of scale and scope as well as building confidence regions for functionals on the identified set, such as the level of the function and its derivative at a point. The obtained procedure is also applicable to a wider class of models defined by a conditional moment restriction, as in Newey & Powell (2003) and Ai & Chen (2003).

KEYWORDS: Bootstrap, instrumental variables, conditional moment, partial identification.

1

1 Introduction

A vast number of estimation problems in economics do not fit the classical regression framework, but are instead of the form:

$$Y = \theta_0(W) + \epsilon \tag{1}$$

where $E[\epsilon|W] \neq 0$, and hence the regressor W is considered to be endogenous. In both theoretical and empirical work, the analysis of (1) has proceeded by employing an instrument Z satisfying the exogeneity condition $E[\epsilon|Z] = 0$. The unknown structural parameter $\theta_0(w)$ can then be characterized as one of the solutions to the integral equation:

$$E[Y - \theta(W)|Z] = 0 \tag{2}$$

Therefore, the model is nonparametrically identified if and only if the solution to (2) is unique. As originally discussed in Newey & Powell (2003), however, said requirement necessitates an instrument satisfying conditions far stronger than the usual covariance restrictions of the parametric model.

Instead of imposing the strong assumptions required for identification, it is possible to study nonparametric instrumental variables as a partially identified model; see Manski (1990, 2003). Hence, rather than requiring a unique solution to equation (2), we examine the *set* of solutions. Referred to as the *identified set*, we denote these solutions by:

$$\Theta_0^{\mathrm{IV}} = \{\theta(w) \in \Theta : E[Y - \theta(W)|Z] = 0\} \tag{3}$$

Santos (2007) develops a family of test statistics for the null hypothesis that a conjectured restriction is satisfied by at least one of the elements in Θ_0^{IV}. These tests can be employed to build confidence regions for functionals of identifiable parameters, as proposed in Romano & Shaikh (2006) elaborating on Imbens & Manski (2004). Such confidence regions $C_n(1-\alpha)$ for functionals $f(\theta) : \Theta_0^{\mathrm{IV}} \to \Re^m$ satisfy the coverage requirement:

$$\inf_{\theta(w) \in \Theta_0^{\mathrm{IV}}} \lim_{n \to \infty} P(f(\theta) \in C_n(1 - \alpha)) = 1 - \alpha \tag{4}$$

This procedure is applicable to a wide range of functionals including the level of a function and its derivatives at a point. We emphasize that if the model is actually identified, then the hypothesis test reduces to whether the true model satisfies the conjectured restriction. Confidence intervals such as (4) are then for the functional applied to $\theta_0(w)$.

The contribution of the present paper is to establish the almost sure consistency of a bootstrap procedure for the family of test statistics proposed in Santos (2007). When identification is not

attained, the limiting distribution of said statistics is nonstandard and thus the bootstrap provides a valuable tool for performing inference. The bootstrap method derived in this paper actually allows for inference in a larger class of models than (1). In particular, it permits inference on the set of solutions to models of the form:

$$E[m(X, \theta)|Z] = 0 \qquad (5)$$

where $m(x, \theta) : \Re^{d_x} \times \Theta \to \Re$ is known and $\theta(x)$ belongs to some nonparametric set Θ. Under identification, these models have been studied by Newey & Powell (2003) who establish consistency of a nonparametric estimator, and Ai & Chen (2003) who construct efficient asymptotically normal estimators for semiparametric specifications.

In related work, Bugni (2008) and Canay (2008) establish the consistency of bootstrap procedures for parametric partially identified models defined by moment inequalities. To the best of our knowledge, the present bootstrap results are the first to be applicable to inference in nonparametric partially identified models. Additional work on nonparametric instrumental variables under the assumption of identification includes Newey, Powell & Vella (1999), Darolles, Florens & Renault (2003), Blundell, Chen & Kristensen (2004), Hall & Horowitz (2005), Horowitz (2006, 2007) and Gagliardini & Scaillet (2007a, 2007b). These models have also been studied without requiring identification in Imbens & Newey (2006), Severini & Tripathi (2006, 2007) Schennach, Chalak & White (2007) and Santos (2008).

The remainder of the paper is organized as follows. Section 2 reviews the test statistic in Santos (2007), as the design of the bootstrap procedure is particular to that setting. Section 3 develops the bootstrap methodology and establishes its almost sure consistency. Section 4 briefly concludes. All proofs are contained in the appendix.

2 The Test Statistic

This section introduces the general testing framework and briefly reviews the principal result in Santos (2007), which will be necessary for the development of the bootstrap procedure.

2.1 General Setup

In order to establish the consistency of the bootstrap, it will be important to ensure certain statistics behave uniformly over the parameter space Θ. For this reason we require Θ to be compact.

Compactness can be attained by imposing bounds on the higher order derivatives of the functions under consideration, see for example Gallant & Nychka (1987). Let $X \in \Re^{d_x}$ and define λ to be a d_x dimensional vector of nonnegative integers. In addition, define $|\lambda| = \sum_{i=1}^{d_x} \lambda_i$ and let $D^\lambda \theta(x) = \partial^{|\lambda|} \theta(x) / \partial x_1^{\lambda_1} \dots \partial x_{d_x}^{\lambda_{d_x}}$. For m, m_0 and δ_0 positive integers satisfying $m > d_x/2$, $\delta_0 > d_x/2$, $(d_x/m_0 + d_x/\delta_0) < 1/2$ and $d_x/2 < \delta < \delta_0$ define the norms:

$$||\theta||_s = \left\{ \sum_{|\lambda| \leq m+m_0} \int \left[D^\lambda \theta(x) \right]^2 (1 + x'x)^{\delta_0} dx \right\}^{\frac{1}{2}} \qquad ||\theta||_{c\delta} = \max_{|\lambda| \leq m_0} \sup_x |D^\lambda \theta(x)|(1 + x'x)^\delta \qquad (6)$$

We will restrict the parameter space to be the set of functions that are bounded in the norm $||\cdot||_s$,

$$\Theta = \{\theta(x) : ||\theta||_s \leq B\} \qquad (7)$$

As noted by Gallant & Nychka (1987), all $\theta(x) \in \Theta$ are also uniformly bounded under $||\cdot||_{c\delta}$. Hence, all such functions have derivatives up to order m_0 uniformly bounded. In our analysis we will not consider *all* solutions to model (5), but instead only those solutions that are sufficiently smooth. Accordingly, we define the identified set to be:

$$\Theta_0 = \{\theta(x) \in \Theta : E[m(X, \theta)|Z] = 0\} \qquad (8)$$

The type of hypothesis test we will focus on, is of whether at least one element of Θ_0 satisfies a conjectured restriction. Formally, for any set of functions R that is closed under $||\cdot||_{c\delta}$ we will be able to test:

$$H_0 : \Theta_0 \cap R \neq \emptyset \qquad H_1 : \Theta_0 \cap R = \emptyset \qquad (9)$$

Because the norm $||\cdot||_{c\delta}$ is very strong, R being closed under it is a weak requirement. For example, R can be the set of weak monotonic or concave functions, as well as the set of production functions reflecting economies of scale or scope. Through test inversion, hypotheses like (9) can also be used to construct confidence regions for functionals that are continuous under $||\cdot||_{c\delta}$, such as the value of a function and its derivative at a point. See Santos (2007) for a detailed discussion.

2.2 Testing Strategy and Assumptions

Under the definition of Θ and the requirement that R be closed under $||\cdot||_{c\delta}$, it is possible to show that $\Theta_0 \cap R \neq \emptyset$ if and only if:

$$\min_{\theta(x) \in \Theta \cap R} E\left[m(X, \theta)E[m(X, \theta)|Z]f_Z(Z)\right] = 0 \qquad (10)$$

4

Employing the characterization in (10) is considerably easier than utilizing (9), because the former does not depend on Θ_0, which is of course unknown. The construction of the test statistic for (9) then proceeds in two steps.

I. Fix $\theta(x) \in \Theta$ and derive a test statistic $T_n(\theta)$ for the null hypothesis $H_0 : E[m(X, \theta)|Z] = 0$, or equivalently, for the null hypothesis $H_0 : E[m(X, \theta)E[m(X, \theta)|Z]f_Z(Z)] = 0$.

II. Let Θ_j be a sieve approximating Θ. Then, following (10), test $H_0 : \Theta_0 \cap R \neq \emptyset$ by using the statistic $I_n(R) = \min_{\theta(x) \in \Theta_j \cap R} T_n(\theta)$.

The intuition behind this procedure is straightforward. Once $\theta(x) \in \Theta$ has been fixed, testing whether $\theta(x) \in \Theta_0$ is equivalent to the nonparametric specification test $H_0 : E[m(X, \theta)|Z] = 0$. This is a well studied testing problem with a variety of available test statistics that have a known asymptotic distribution if $\theta(x) \in \Theta_0$, but diverge to infinity if $\theta(x) \notin \Theta_0$. If $\Theta_0 \cap R = \emptyset$, then when computing $I_n(R)$ we will minimize $T_n(\theta)$ over values for which it diverges to infinity. Hence, $I_n(R)$ will diverge to infinity as well. On the other hand, if $\Theta_0 \cap R \neq \emptyset$, then the minimum value of $T_n(\theta)$ over $\Theta \cap R$ will be attained in a neighborhood of $\Theta_0 \cap R$, as $T_n(\theta)$ diverges to infinity for all other values. Utilizing the limiting distribution of $T_n(\theta)$ when $\theta(x) \in \Theta_0$, it is then possible to find the asymptotic distribution of $I_n(R)$.

For Step I, testing $H_0 : E[m(X, \theta)|Z] = 0$ for a fixed $\theta(x) \in \Theta$, we employ the following test-statistic studied in Zheng (1996):

$$T_n(\theta) = \frac{2}{(n-1)h^{\frac{d_z}{2}}} \sum_{i=2}^{n} \sum_{j<i} K\left(\frac{z_i - z_j}{h}\right) m(x_i, \theta)m(x_j, \theta) \tag{11}$$

where $Z \in \Re^{d_z}$, $K(\cdot)$ is a kernel with properties to be specified in the assumptions, and h is the bandwidth. Accordingly, for Step II we define the test statistic:

$$I_n(R) = \min_{\theta(x) \in \Theta_j \cap R} T_n(\theta) \tag{12}$$

When computing $I_n(R)$ we employ a parametric sieve Θ_j because minimizing over the nonparametric set of functions $\Theta \cap R$ might prove unfeasible.

The following assumptions are sufficient for obtaining the asymptotic distribution of $I_n(R)$.

ASSUMPTION 1: (i) $\{x_i, z_i\}_{i=1}^{n}$ are i.i.d. with $X \in \Re^{d_x}$ and $Z \in \Re^{d_z}$; (ii) Z is continuously distributed with density $f_Z(z)$; (iii) $f_Z(z)$ is continuous and bounded; (iv) R is closed under $||\cdot||_{c\delta}$.

ASSUMPTION 2: (i) All derivatives up to order k of $(E[m(X, \theta)|z])^j f_Z(z)$ are bounded uniformly in $\theta(x) \in \Theta$ and $1 \leq j \leq 2$; (ii) The functions $m(x, \theta)$ have envelope $F(x)$ with $E[F^j(X)|z]$ bounded

and continuous for $1 \leq j \leq \bar{u}$; (iii) For every $\theta_1(x), \theta_2(x) \in \Theta$, $|m(x, \theta_1) - m(x, \theta_2)| \leq G(x)||\theta_1 - \theta_2||_\infty$ with $E[G^j(X)|z]$ bounded and continuous for $1 \leq j \leq \bar{u}$; (iv) $\bar{u} \geq 4$.

ASSUMPTION 3: (i) The kernel function $K(\cdot) : \Re^{d_z} \to \Re$ is symmetric of order $\lfloor k \rfloor - 1$; (ii) The bandwidth satisfies $h \to 0$ and $nh^{d_z} \to \infty$ and $nh^{\frac{d_z}{2} + \lfloor k \rfloor} \to 0$; (iii) For some $l \leq \lfloor k \rfloor$, $\sqrt{n}h^l \to \infty$ and $nh^{d_z + l\left(1 - \frac{(m_0 + \delta_0)d_x}{2m_0\delta_0}\right)} \to 0$; (iv) The sieve $\{\Theta_j\} \subseteq \Theta$ are closed under the norm $|| \cdot ||_\infty$ and $\sup_{\Theta \cap R} \inf_{\Theta_j \cap R} ||\theta - \theta_j||_\infty = o((nh^{\frac{d_z}{2}})^{-1})$.

Assumptions 1(i)-(iii) describe the sampling process, while Assumption 1(iv) is necessary to ensure that (10) holds if and only if $\Theta_0 \cap R \neq \emptyset$. Assumptions 2(ii)-(iv) specify the regularity conditions for $m(x, \theta)$. Assumption 2(ii) and 2(iv) implies moments up to $\bar{u}$ of $\sup_\Theta |m(X, \theta)|$ exist, while the Lipschitz requirement in Assumption 2(iii) enables us to control the uniform behavior of the empirical process over the class of functions $\mathcal{F} = \{m(x, \theta) : \theta \in \Theta\}$. Assumptions 2(i) and 3(i)-(iii) in turn guarantee that the biases present in nonparametric estimation vanish at the correct rates. Assumption 3(iv) requires the sieve $\{\Theta_j \cap R\}$ to be able to approximate $\Theta \cap R$ uniformly well under $|| \cdot ||_\infty$. For testing global restrictions such as monotonicity or concavity, particular sieves such as shape preserving splines need to be employed. See Chen (2006) for examples.

2.3 Asymptotic Distribution

We now describe the asymptotic distribution of the statistics $T_n(\theta)$ and $I_n(R)$, defined in (11) and (12) respectively. Zheng (1996) proposes a consistent test of functional form in nonlinear regression models. Once $\theta(x) \in \Theta$ is fixed, testing the null hypothesis $H_0 : E[m(X, \theta)|Z] = 0$ is then a special case of the testing problem in Zheng (1996). The proof of Lemma 2.1, therefore follows readily from his results.

Lemma 2.1. *If Assumptions 1(i)-(ii), 2(ii) and 3(i)-(ii) hold and $\theta(x) \in \Theta_0$, then*

$$T_n(\theta) \xrightarrow{\mathcal{L}} N(0, \sigma^2(\theta))$$

where $\sigma^2(\theta) = 2\left[\int K^2(u)du\right] E\left[(E[m^2(X, \theta)|Z])^2 f_Z(Z)\right]$. Furthermore, if $\theta(x) \notin \Theta_0$, then

$$T_n(\theta) \xrightarrow{p} +\infty$$

The assumptions used to establish Lemma 2.1 are stronger than those in Zheng (1996). They are, however, necessary to study the uniform behavior of $T_n(\theta)$ on Θ rather than at just one particular function $\theta(x) \in \Theta$. These assumptions allow us to establish Theorem 2.1, which is the main result in Santos (2007).

Theorem 2.1. *If Assumptions 1(i)-(iv), 2(i)-(iv) and 3(i)-(iv) hold and $\Theta_0 \cap R \neq \emptyset$, then*

$$I_n(R) \xrightarrow{\mathcal{L}} \min_{\theta(x) \in \Theta_0 \cap R} G(\theta)$$

where $G(\theta)$ is a Gaussian process in $\mathcal{L}^\infty(\Theta_0)$. Under the same assumptions, if $\Theta_0 \cap R = \emptyset$, then

$$I_n(R) \xrightarrow{p} \infty.$$

The space $\mathcal{L}^\infty(\Theta_0)$ consists of all bounded functionals $f(\theta) : \Theta_0 \to \Re$. See Chapter 1.5 in van der Vaart and Wellner (1996) for a detailed discussion of stochastic processes defined on these spaces. If the model is identified, then Θ_0 is a singleton. The Gaussian process $G(\theta)$ from Theorem 2.1 is consequently defined at a single point and hence it is just a univariate normally distributed random variable. On the other hand, when the model is not identified and $\Theta_0 \cap R$ is not a singleton, Theorem 2.1 implies the asymptotic distribution is nonstandard. For this reason the bootstrap results of the present paper are very useful for inference.

3 Bootstrap Consistency

Li & Wang (1998) derive the consistency in probability of a bootstrap procedure for the test-statistic of Zheng (1996). We build off their results to show the almost sure consistency of a wild bootstrap for the statistic $T_n(\theta)$ that works uniformly in Θ. This result can then be employed to obtain an almost sure consistent bootstrap for the statistic $I_n(R)$.

Throughout this section we will denote $w_i = (z_i, x_i)$ and let $\mathcal{W}_n$ be the σ-field generated by $\{w_i\}_{i=1}^n$. We also define $\mathcal{L}^*$ to be the law of a random variable conditional on $\mathcal{W}_n$ and denote for any random variable V, $E^*[V] = E[V|\mathcal{W}_n]$. All bootstrap consistency results obtained in this paper are almost surely in the possible sample realizations $\{w_i\}_{i=1}^n$.

3.1 Bootstrap Procedure for $T_n(\theta)$

If $\theta(x) \in \Theta_0$, then $E[m(X, \theta)|Z] = 0$ implies that the U-Statistic $T_n(\theta)$ is degenerate of order one. In contrast, when $\theta(x) \notin \Theta_0$, $T_n(\theta)$ is no longer degenerate for the bandwidth h sufficiently small. Regular bootstrap procedures are often inconsistent for degenerate U-Statistics. Arcones & Gine (1992) show that the key to restoring bootstrap consistency is to ensure the degeneracy of the bootstrap statistic conditional on the sample.

Li & Wang (1998) impose degeneracy through the wild bootstrap. In particular, define a random variable U with the following distribution:

$$P\left(U = -\frac{\sqrt{5}-1}{2}\right) = \frac{\sqrt{5}+1}{2\sqrt{5}} \qquad\qquad P\left(U = \frac{\sqrt{5}+1}{2}\right) = \frac{\sqrt{5}-1}{2\sqrt{5}} \qquad (13)$$

To implement the wild bootstrap, we generate a sample $\{u_i\}_{i=1}^n$ distributed according to (13) and independent of $\mathcal{W}_n$. Employing the sample $\{u_i\}_{i=1}^n$ we then examine the bootstrap statistic:

$$T_n^*(\theta) = \frac{2h^{\frac{d_z}{2}}}{n-1} \sum_{i=2}^n \sum_{j<i} K\left(\frac{z_i - z_j}{h}\right) m^*(x_i,\theta)m^*(x_j,\theta) \qquad m^*(x_i,\theta) = m(x_i,\theta)u_i \qquad (14)$$

Under Assumptions stronger than Li & Wang (1998) we modify their result to show the almost surely consistency of the bootstrap instead of in probability.

Lemma 3.1. *If Assumptions 1(i)-(iii), 2(ii)-(iii) and 3(i)-(ii) hold with $\bar{u} \geq 10$, then it follows that for every $\theta(x) \in \Theta$,*

$$T_n^*(\theta) \xrightarrow{\mathcal{L}^*} N(0, \sigma^2(\theta)) \quad a.s.$$

where $\sigma^2(\theta) = 2\left[\int K^2(u)du\right] E\left[(E[m^2(X,\theta)|Z])^2 f_Z(Z)\right].$

Lemma 3.1 is an inmediate consequence of the properties of the multiplier random variable U. Since $E[U] = 0$ and the sample $\{u_i\}_{i=1}^n$ is i.i.d. and independent of $\mathcal{W}_n$, it follows that $T_n^*(\theta)$ is a degenerate U-Statistic of order one under $\mathcal{L}^*$. A martingale central limit theorem, as in Hall (1984), enables us to establish the asymptotic normality of $T_n^*(\theta)$ under $\mathcal{L}^*$. Furthermore, since $E[U^2] = 1$, the asymptotic variance of $T_n^*(\theta)$ is the same as that of $T_n(\theta)$ when $\theta(x) \in \Theta_0$. Thus, as both $T_n(\theta)$ and $T_n^*(\theta)$ are asymptotically normally distributed with the same variance under their respective laws, the consistency of the bootstrap is implied. This result is analogous to the multiplier central limit theorem, see Chapter 2.9 in van der Vaart & Wellner (1996).

3.2 Bootstrap Procedure for $I_n(R)$

The test statistic $T_n(\theta)$ is degenerate of order one if $\theta(x) \in \Theta_0$ and is non-degenerate otherwise. As a result, $T_n(\theta)$ diverges to infinity when $\theta(x) \notin \Theta_0$ because in this case $T_n(\theta)$ is not only improperly centered but its variance diverges to infinity as well. In contrast, the wild bootstrap procedure imposes degeneracy on $T_n^*(\theta)$. Therefore, the bootstrap statistic $T_n^*(\theta)$ is degenerate of order one under $\mathcal{L}^*$ for *all* $\theta(x) \in \Theta$. As a result, Lemma 3.1 establishes $T_n^*(\theta)$ is conditionally asymptotically normally distributed for *all* $\theta(x) \in \Theta$ instead of only for $\theta(x) \in \Theta_0$.

In Lemma 3.2 we refine the result of Lemma 3.1. The bootstrap statistic $T_n^*(\theta)$ not only converges in law to a normal distribution, but it does so uniformly in $\theta(x) \in \Theta$. Considered as a *process* on $\mathcal{L}^\infty(\Theta)$, the bootstrap statistic $T_n^*(\theta)$ converges in law to a Gaussian process. Furthermore, the limiting Gaussian process has the same covariance structure on $\mathcal{L}^\infty(\Theta_0)$ as $G(\theta)$ from Theorem 2.1.

Lemma 3.2. *If Assumptions 1(i)-(iii), 2(ii)-(iii) and 3(i)-(ii) hold with $\bar{u} \geq 10$, then*

$$T_n^*(\theta) \xrightarrow{\mathcal{L}^*} G^b(\theta) \quad a.s.$$

where $G^b(\theta)$ is a Gaussian process on $\mathcal{L}^\infty(\Theta)$ that agrees with $G(\theta)$ from Theorem 2.1 on $\mathcal{L}^\infty(\Theta_0)$.

While $T_n^*(\theta)$ converges in law to the correct process on $\mathcal{L}^\infty(\Theta_0)$, the fact that it also converges in law for all $\theta(x) \notin \Theta_0$ presents a problem. The asymptotic distribution of the original test statistic $I_n(R)$ depends crucially on the fact that $T_n(\theta)$ diverges to infinity for all $\theta(x) \notin \Theta_0$. When calculating $I_n(R)$, the minimum has to be attained in a shrinking neighborhood of Θ_0 since $T_n(\theta)$ is only well behaved for $\theta(x) \in \Theta_0$. Unfortunately, because $T_n^*(\theta)$ is asymptotically normally distributed for all $\theta(x) \in \Theta$, the above argument no longer applies. It is possible to show through the continuous mapping theorem that the direct analogue to $I_n(R)$ has the following asymptotic distribution:

$$\min_{\theta(x) \in \Theta_j \cap R} T_n^*(\theta) \xrightarrow{\mathcal{L}^*} \min_{\theta(x) \in \Theta \cap R} G^b(\theta) \quad a.s. \tag{15}$$

where $G^b(\theta)$ is the Gaussian process in Lemma 3.2. Because $\Theta_0 \subseteq \Theta$ and $G^b(\theta)$ agrees with $G(\theta)$ on $\mathcal{L}^\infty(\Theta_0)$, calculating critical values from the bootstrap distribution of $\min_{\theta(x) \in \Theta_j \cap R} T_n^*(\theta)$ would yield a potentially severely conservative procedure.

In order to remedy this problem, we require an indicator for whether $\theta(x) \in \Theta_0$ or not. Asymptotically we can then examine the minimum of the process $T_n^*(\theta)$ over $\Theta_0 \cap R$ only instead of $\Theta \cap R$. A natural candidate for such indicator is $T_n(\theta)$ itself. In particular we will employ

$$n^{-r} T_n(\theta)$$

for $0 < r < 1/2$. When $\theta(x) \in \Theta_0$, $T_n(\theta)$ is asymptotically normally distributed. Hence, under the appropriate moment conditions, setting $r > 0$ implies $n^{-r} T_n(\theta) \xrightarrow{as} 0$. On the other hand, if $\theta(x) \notin \Theta_0$, then $T_n(\theta)$ is no longer degenerate and its standard deviation is of order $O(\sqrt{n})$. Therefore, if $r < 1/2$, then under the right moment conditions $n^{-r} T_n(\theta) \xrightarrow{as} \infty$. Given these properties of $n^{-r} T_n(\theta)$, we define the relevant bootstrap statistic for $I_n(R)$ to be:

$$I_n^*(R) = \min_{\theta(x) \in \Theta_j \cap R} \left(n^{-r} T_n(\theta) + T_n^*(\theta) \right) \tag{16}$$

9

Intuitively, since $n^{-r}T_n(\theta)$ diverges to infinity for all $\theta(x) \notin \Theta_0$, the minimum in (16) must be attained in a shrinking neighborhood of $\Theta_0 \cap R$. In turn, since the limiting law of the process $T_n^*(\theta)$ agrees with that of $T_n(\theta)$ on $\mathcal{L}^\infty(\Theta_0)$, the law of $I_n^*(R)$ is almost sure consistent for that of $I_n(R)$. Assumption 4 is sufficient for formalizing this argument.

ASSUMPTION 4: (i) $(\bar{u}, r)$ satisfy $\bar{u} \geq 10$, $2r\lfloor \frac{\bar{u}}{2} \rfloor > 1$ and $n^{\gamma + (1-2r)\frac{\bar{u}}{2}} h^{\left(d_z + l - \frac{l(m_0 + \delta_0)d_x}{2m_0\delta_0} \right)\frac{\bar{u}}{2}} \to 0$ for some $\gamma > 1$ and $l \leq \lfloor k \rfloor$ such that $n^{\frac{1}{2}-r} h^l \to \infty$; (ii) $\{\Theta_j\} \subseteq \Theta$ are closed under $||\cdot||_\infty$ and $\sup_{\Theta \cap R} \inf_{\Theta_j \cap R} ||\theta - \theta_j||_\infty = o((n^{1-r}h^{\frac{d_z}{2}})^{-1})$.

The parameter $\bar{u}$, defined in Assumption 2(ii), allows us to control the tails of the process $T_n(\theta)$ and thus influences what values of r are permissible in (16). Given these Assumptions we establish the main result of this paper.

Theorem 3.1. *If Assumptions 1(i)-(iv), 2(i)-(iii), 3(i)-(iii), 4(i)-(ii) hold and $\Theta_0 \cap R \neq \emptyset$, then*

$$I_n^*(R) \xrightarrow{\mathcal{L}^*} \inf_{\theta(x) \in \Theta_0 \cap R} G(\theta) \quad a.s.$$

where $G(\theta)$ is a Gaussian process on $\mathcal{L}^\infty(\Theta_0)$ with the same marginals as in Theorem 2.1. Furthermore, under the same assumptions, if $\Theta_0 \cap R = \emptyset$, then

$$I_n^*(R) \xrightarrow{p^*} \infty \quad a.s.$$

In Section 3.3 we conclude the theoretical derivations by showing how to use Theorem 3.1 to obtain the appropriate critical values necessary for inference.

3.3 Implementation of Bootstrap Inference

Theorem 2.1 establishes that under the null hypothesis $H_0 : \Theta_0 \cap R \neq \emptyset$, the asymptotic distribution of our test statistic $I_n(R)$ is given by:

$$I_n(R) \xrightarrow{\mathcal{L}} \min_{\theta(x) \in \Theta_0 \cap R} G(\theta) \tag{17}$$

On the other hand, if $\Theta_0 \cap R = \emptyset$, then $I_n(R) \xrightarrow{p} \infty$. Therefore, if the quantiles of (17) were known it would be straightforward to construct a consistent test for $H_0 : \Theta_0 \cap R$ with a desired size. In particular, define:

$$c_{1-\alpha} = \inf \left\{ x : P\left(\min_{\theta(x) \in \Theta_0 \cap R} G(\theta) \leq x \right) \geq 1 - \alpha \right\} \tag{18}$$

It is then immediate from Theorem 2.1, that a test that rejects whenever $I_n(R) > c_{1-\alpha}$ is both consistent and has size α. While $c_{1-\alpha}$ is unknown, we can estimate it by employing the analogous

quantile from the bootstrap statistic $I_n^*(R)$. Hence, let:

$$\hat{c}_{1-\alpha} = \inf \left\{ x : P^* \left(I_n^*(R) \le x \right) \ge 1 - \alpha \right\} \tag{19}$$

The bootstrap quantile can be obtained analytically since the probability law $\mathcal{L}^*$ is, conditional on the sample, generated according to (13). Alternatively, $\hat{c}_{1-\alpha}$ can be easily obtained through simulation, as outlined in Steps 1-3.

STEP 1: Generate a sample $\{u_i\}_{i=1}^n$ with U distributed according to (13).

STEP 2: Use the sample $\{u_i\}_{i=1}^n$ to compute $I_n^*(R)$.

STEP 3: Repeat Steps 1-2 S times to obtain S statistics $I_n^*(R)$. The S sample $1 - \alpha$ quantile is then consistent for $\hat{c}_{1-\alpha}$ as the number of simulations S goes to infinity.

In Theorem 3.2 we establish that conducting the test using the critical value $\hat{c}_{1-\alpha}$ instead of $c_{1-\alpha}$ still allows us to control the size and remain consistent.

Theorem 3.2. *If Assumptions 1(i)-(iv), 2(i)-(iii), 3(i)-(iii), 4(i)-(ii) hold and $\Theta_0 \cap R \ne \emptyset$, then*

$$\lim_{n\to\infty} P \left(I_n(R) \le \hat{c}_{1-\alpha} \right) = 1 - \alpha$$

Furthermore, under the same assumptions, if $\Theta_0 \cap R = \emptyset$, then it follows that

$$\lim_{n\to\infty} P \left(I_n(R) > \hat{c}_{1-\alpha} \right) = 1$$

The ability to control the size by employing $\hat{c}_{1-\alpha}$ follows readily from the almost sure consistency of the bootstrap, see for example Beran (1984). The consistency of a test based on the bootstrap is due to the different rates at which $I_n^*(R)$ and $I_n(R)$ diverge to infinity. When $\Theta_0 \cap R = \emptyset$, it follows from Theorems 2.1 and 3.1 that:

$$I_n^*(R) \equiv \inf_{\Theta_j \cap R} \left(n^{-r} T_n(\theta) + T_n^*(\theta) \right) \xrightarrow{p^*} \infty \quad a.s. \qquad I_n(R) \equiv \inf_{\Theta_j \cap R} T_n(\theta) \xrightarrow{p} \infty \tag{20}$$

By Lemma 3.2, however, $T_n^*(\theta)$ converges in law to a tight Gaussian process almost surely. Therefore, the divergence of $I_n^*(R)$ is exclusively caused by $n^{-r} T_n(\theta)$ diverging to infinity uniformly on $\Theta \cap R$. Because $I_n(R)$ diverges to infinity due to $T_n(\theta)$ as well, $r > 0$ implies $I_n^*(R)$ eventually falls behind $I_n(R)$. In particular, we show in the appendix that:

$$P^* \left(I_n^*(R) < I_n(R) \right) \xrightarrow{as} 1 \,, \tag{21}$$

which establishes the consistency of the test based on the bootstrap critical value $\hat{c}_{1-\alpha}$.

4   Conclusion

Nonparametric identification of instrumental variables models can be hard to attain. In this cases it is still possible to perform inference on the identified set. In Santos (2007) we derive a test statistic for the null hypothesis that at least one element of the identified set satisfies a conjectured restriction. Without identification, however, the asymptotic distribution of the test statistic is nonstandard. The present paper addresses this problem by proposing a bootstrap procedure and establishing its almost sure consistency. This procedure is also applicable to a wider class of models including Newey & Powell (2003) and Ai & Chen (2003). Even if the model is identified, the bootstrap may provide a higher order refinement than an asymptotic approximation. This question is beyond the scope of this paper, as its demonstration requires obtaining an Edgeworth expansion for our test statistic. We plan to address this problem in future work.

The proofs of Lemma 2.1 and Theorem 2.1 are omitted from the Appendix and can be found in Santos (2007). Throughout the Appendix we will let $w_i = (x_i, z_i)$ and define the kernels

$$H_n(w_i, w_j, \theta) = h^{-d_z} K\left(\frac{z_i - z_j}{h}\right) m(x_i, \theta) m(x_j, \theta) \qquad H_n^*(w_i, w_j, \theta) = h^{-d_z} K\left(\frac{z_i - z_j}{h}\right) m^*(x_i, \theta) m^*(x_j, \theta)$$

Notice that with these definitions, we have

$$T_n(\theta) = \frac{2h^{\frac{d_z}{2}}}{(n-1)} \sum_{i=2}^{n} \sum_{j<i} H_n(w_i, w_j, \theta) \qquad T_n^*(\theta) = \frac{2h^{\frac{d_z}{2}}}{(n-1)} \sum_{i=2}^{n} \sum_{j<i} H_n^*(w_i, w_j, \theta)$$

In the following table we include additional notation and definitions that will be introduced and used throughout the appendix, including many that go beyond the ones already introduced in the main text.

$a \lesssim b$	$a \leq Mb$ for some constant M which is universal in the context of the proof
$\|\theta\|_\infty$	The sup-norm $\sup_x \|\theta(x)\|$
$\|\theta\|_s$	The norm $\left\{\sum_{\|\lambda\| \leq m+m_0} \int \left[D^\lambda \theta(x)\right]^2 (1 + x'x)^{\delta_0} dx\right\}^{\frac{1}{2}}$ with $m > k/2$, $\delta_0 > k/2$ and $(k/m_0 + k/\delta_0) < 1/2$
$\|\theta\|_{c\delta}$	The norm $\max_{\|\lambda\| \leq m_0} \sup_x \|D^\lambda \theta(x)\|(1 + x'x)^\delta$ for $k/2 < \delta < \delta_0$
$\|\theta\|_{\mathcal{L}^2(X)}$	The norm $\left(E[\theta^2(X)]\right)^{\frac{1}{2}}$
$\rho_{mn}(f, g)$	The random semimetric $\left[n^{-m} \sum_{I_m^N} (f(x_{i_1}, \ldots, x_{i_m}) - g(x_{i_1}, \ldots, x_{i_m}))^2\right]^{\frac{1}{2}}$
$\|f\|_{mn}$	The random norm $\left[n^{-m} \sum_{I_m^n} f^2(x_{i_1}, \ldots, x_{i_m})\right]^{\frac{1}{2}}$
$H_n(w_i, w_j, \theta)$	The U-Statistic Kernel $h^{-d_z} K\left(\frac{z_i - z_j}{h}\right) m(x_i, \theta) m(x_j, \theta)$.
$H_n^*(w_i, w_j, \theta)$	The U-Statistic Kernel $h^{-d_z} K\left(\frac{z_i - z_j}{h}\right) m^*(x_i, \theta) m^*(x_j, \theta)$.
I_m^n	Set of distinct m-tuples from N observations
$m^*(x_i, \theta)$	The function $m(x_i, \theta) u_i$ where u_i is distributed according to (13).
$\mathcal{L}^\infty(\Theta)$	The metric space of bounded functionals on Θ with norm $\sup_\Theta \|F_1(\theta) - F_2(\theta)\|$
$N(\mathcal{F}, \|\cdot\|, \epsilon)$	Covering numbers of size ϵ for $\mathcal{F}$ under the norm $\|\cdot\|$
$N_{[]}(\mathcal{F}, \|\cdot\|, \epsilon)$	Bracketing numbers of size ϵ for $\mathcal{F}$ under the norm $\|\cdot\|$
P-Canonical function	A function $f(X^m)$ satisfying $E[f(X^m)|X^n] = 0$ for all $n < m$
$T_n(\theta)$	The U-Statistic $\frac{2h^{\frac{d_z}{2}}}{n-1} \sum_{i=2}^{n} \sum_{j<i} H_n(w_i, w_j, \theta)$.
$T_n^*(\theta)$	The U-Statistic $\frac{2h^{\frac{d_z}{2}}}{n-1} \sum_{i=2}^{n} \sum_{j<i} H_n^*(w_i, w_j, \theta)$.
X^m	The random variable consisting of m independent copies of the random variable X
w_n	The random variables (x_n, z_n)

APPENDIX B - Proof of Lemma 3.1 and Auxiliary Lemmas .1 and .2

Proof of Lemma 3.1: The proof follows a martingale argument as in Hall (1984). First we define,

$$W_{ij} \equiv K\left(\frac{z_i - z_j}{h}\right) \qquad Y_i^* \equiv m^*(x_i, \theta) \sum_{j=1}^{i-1} W_{ij} m^*(x_j, \theta) \tag{22}$$

Next, notice that with these definitions, the equality in (23) immediately follows.

$$T_n^*(\theta) = \frac{2h^{-\frac{d_z}{2}}}{(n-1)} \sum_{i=2}^{n} Y_i^* \tag{23}$$

Let $\mathcal{F}_i$ denote the σ-field generated by $\{m^*(X_1, \theta), \ldots, m^*(X_i, \theta)\}$, conditional on the sample, for $0 \leq i \leq n$. Then note that, $E^*[Y_i^* | \mathcal{F}_{i-1}] = 0$ for all i. Therefore, $\{S_n = \sum_{i=1}^n Y_i^*, \mathcal{F}_i\}$ is a martingale conditional on the sample. We aim to apply a central limit theorem for martingales. In (24) we derive the conditional variance of S_n, using $E^*[(m^*(x_i, \theta))^2] = m^2(x_i, \theta)$.

$$
\begin{aligned}
V_n^* &\equiv \sum_{i=2}^n E^*[(Y_i^*)^2 | \mathcal{F}_i] = \sum_{i=2}^n m^2(x_i, \theta) \left[\sum_{j=1}^{i-1} W_{ij} m^*(x_j, \theta) \right]^2 \\
&= \sum_{i=2}^n m^2(x_i, \theta) \sum_{j=1}^{i-1} (m^*(x_j, \theta))^2 W_{ij}^2 + 2 \sum_{i=3}^n m^2(x_i, \theta) \sum_{j=2}^{i-1} \sum_{k<j} W_{ij} W_{ik} m^*(x_j, \theta) m^*(x_k, \theta) \\
&\equiv V_{n1}^* + V_{n2}^*
\end{aligned}
\tag{24}
$$

Next, notice that by Lemma .1 result (25) immediately follows.

$$
n^{-2} h^{-d_z} V_n^* \xrightarrow{p^*} \sigma^2(\theta)/2 \quad a.s.
\tag{25}
$$

In addition, Lemma .2 verifies a Lindenberg-Feller type condition for the martingale $\{S_n = \sum_{i=1}^n Y_i^*, \mathcal{F}_i\}$, and therefore Theorem VIII.1.1 in Pollard (1984) establishes (26).

$$
T_n^*(\theta) \xrightarrow{\mathcal{L}^*} N(0, \sigma^2(\theta)) \quad a.s.
\tag{26}
$$

Which concludes the proof of the Lemma. ∎

Lemma .1. *If Assumptions 1(i)-(iii), 2(ii)-(iii), 3(i)-(ii) with $\bar{u} \geq 8$, then V_{n1}^* and V_{n2}^* as in (24), satisfy*

$$
n^{-2} h^{-d_z} V_{n1}^* \xrightarrow{p^*} \sigma^2(\theta)/2 \quad a.s. \qquad n^{-2} h^{-d_z} V_{n2}^* \xrightarrow{p^*} 0 \quad a.s.
$$

where $\sigma^2(\theta) = 2 \left[\int K^2(u) du \right] E \left[(E[m^2(X, \theta)|Z])^2 f_Z(Z) \right]$.

Proof of Lemma .1: We first study $n^{-2} h^{-d_z} V_{n1}^*$. Define,

$$
H_n^1(w_i, w_j) = W_{ij}^2 m^2(x_i, \theta) m^2(x_j, \theta)
\tag{27}
$$

Because $E^*[(m^*(x_i, \theta))^2] = m^2(x_i, \theta)$, the equality in (28) then follows.

$$
E^*[V_{n1}^*] = \sum_{i=2}^n \sum_{j=1}^{i-1} H_n^1(w_i, w_j) \equiv V_{n1}
\tag{28}
$$

Using V_{n1} as defined in (28), the first equality in (29) then follows by inspection, while the second equality is implied by the change of variables $u = (z_i - z_j)/h$. For the final equality notice the dominated convergence theorem can be applied due to Assumption 2(ii).

$$
\begin{aligned}
n^{-2} h^{-d_z} E[V_{n1}] &= \frac{(n-1) h^{-d_z}}{2n} \int E[m^2(X_i, \theta)|z_i] K^2 \left(\frac{z_i - z_j}{h} \right) E[m^2(X_j, \theta)|z_j] f_Z(z_i) f_Z(z_j) dz_i dz_j \\
&= \frac{(n-1)}{2n} \int E[m^2(X_i, \theta)|z_i] K^2(u) E[m^2(X_i, \theta)|z_i - hu] f_Z(z_i - hu) f_z(z_i) du dz_i \\
&= \frac{1}{2} \left[\int K^2(u) du \right] E \left[(E[m^2(X, \theta)|Z])^2 f_Z(Z) \right] + o(1)
\end{aligned}
\tag{29}
$$

Let $P_n^i(V_1)$ denote the ith term in the Hoeffding decomposition of $2n^{-1}(n-1)^{-1}h^{-d_z}V_{n1}$:

$$P_n^0(V_1) = h^{-d_z}E[H_n^1] \tag{30}$$

$$P_n^1(V_1) = \frac{h^{-d_z}}{n}\sum_{i=1}^{n}E[H_n^1|w_i] - E[H_n^1] \tag{31}$$

$$P_n^2(V_1) = \frac{2h^{-d_z}}{n(n-1)}\sum_{i=2}^{n}\sum_{j<i}H_n^1(w_i,w_j) - E[H_n^1|w_i] - E[H_n^1|w_j] + E[H_n^1] \tag{32}$$

The first equality in (33) then follows by direct calculation. The inequality in (33) is in turn implied by the Cauchy-Schwarz inequality.

$$E\left[\left(P_n^1(V_1)\right)^4\right] = \frac{1}{n^3h^{4d_z}}E\left[\left(E[H_n^1|W_i] - E[H_n^1]\right)^4\right] + \frac{(n-1)}{2n^3h^{4d_z}}E\left[\left(E[H_n^1|W_i] - E[H_n^1]\right)^2\left(E[H_n^1|W_j] - E[H_n^1]\right)^2\right]$$

$$\leq \frac{1}{n^3h^{4d_z}}E\left[\left(E[H_n^1|W_i] - E[H_n^1]\right)^4\right] + \frac{1}{2n^2h^{4d_z}}E\left[\left(E[H_n^1|W_i] - E[H_n^1]\right)^4\right] \tag{33}$$

In (34) we examine $E[(E[H_n^1|W_i])^j]$ for $2j \leq \bar{u}$. The first equality in (34) is implied from (27), and the second equality can be obtained from the change of variables $u = (z_i - z_j)/h$. For the final result use the dominated convergence theorem, which can be applied thanks to Assumption 2(ii).

$$h^{-jd_z}E\left[\left(E[H_n^1|W_i]\right)^j\right] = h^{-jd_z}\int E[m^{2j}(X_i,\theta)|z_i]\left(\int K\left(\frac{z_i-z_j}{h}\right)E[m^2(X_j,\theta)|z_j]f_Z(z_j)dz_j\right)^j f_Z(z_i)dz_i$$

$$= \int E[m^{2j}(X_i,\theta)|z_i]\left(\int K(u)E[m^2(X_i,\theta)|z_i-hu]f_Z(z_i-hu)du\right)^j f_Z(z_i)dz_i$$

$$= E\left[m^{2j}(X,\theta)\left(E[m^2(X,\theta)|Z]\right)^j f_Z^j\right] + o(1) \tag{34}$$

Therefore, combining (33) and (34), we conclude (35).

$$E\left[\left(P_n^1(V_1)\right)^4\right] = O(n^{-2}) \tag{35}$$

Hence, using (35) and Chebychev's inequality we obtain (36).

$$\sum_{n=1}^{\infty}P(|P_n^1(V_1)| > \epsilon) \leq \sum_{n=1}^{\infty}\frac{1}{\epsilon^4}E\left[\left(P_n^1(V_1)\right)^4\right] < \infty \tag{36}$$

In turn, (37) then follows by the Borel-Cantelli Lemma.

$$P_n^1(V_1) \xrightarrow{a.s.} 0 \tag{37}$$

Similarly, in (38), we examine $P_n^2(V_1)$. Notice that the degeneracy of $P_n^2(V_1)$ implies the first equality in (38). The first inequality is standard for Hoeffding decompositions and follows from the kernel of the terms in the Hoeffding decomposition being projections under $||\cdot||_{\mathcal{L}^2}$ of $H_n^1(w_i,w_j)$.

$$E\left[\left(P_n^2(V_1)\right)^2\right] = \frac{4h^{-2d_z}}{(n-1)^2n^2}\sum_{i=2}^{n}\sum_{j=1}^{i-1}E\left[\left(H_n^1(W_i,W_j) - E[H_n^1|W_i] - E[H_n^1|W_j] + E[H_n^1]\right)^2\right]$$

$$\leq \frac{2h^{-2d_z}}{n(n-1)}E\left[\left(H_n^1(W_i,W_j)\right)^2\right] \tag{38}$$

Hence, combining (38) with arguments as in (29) we obtain (39). For the second equality, notice that Assumptions 3(ii)-(iii) imply $n^\delta = o(nh^{d_z})$ for some $\delta > 0$.

$$E\left[\left(P_n^2(V_1)\right)^2\right] = O(n^2h^{d_z}) = O(n^{1+\delta}) \tag{39}$$

15

Hence, arguing as in (36) and (37) it follows that $P_n^2(V_1) \xrightarrow{a.s.} 0$. Combine this result with (28), (29), (37) and $2n^{-1}(n-1)^{-1}h^{-d_z}V_{n1} = \sum_{i=0}^{2} P_n^i(V_1)$ to establish (40).

$$E^*[V_{n1}^*] \xrightarrow{a.s.} \sigma^2(\theta)/2 \tag{40}$$

Next, we study $E^*[(V_{n1}^* - E^*[V_{n1}^*])^2]$. In (41), the first equality follows from (28), the second equality from $E^*[(m^*(x_i,\theta))^4] = 2m^4(x_i,\theta)$ and the third equality by calculation. Denote the resulting terms by $V_{n1}^{(1)}$ and $V_{n1}^{(2)}$.

$$
\begin{aligned}
E^*\left[(V_{n1}^* - E^*[V_{n1}^*])^2\right] &= E^*\left[\left(\sum_{j=1}^{n-1}[(m^*(x_j,\theta))^2 - m^2(x_j,\theta)]\sum_{i=j+1}^{n}m^2(x_i,\theta)W_{ji}^2\right)^2\right] \\
&= \sum_{j=1}^{n-1}m^4(x_j,\theta)\left[\sum_{i=j+1}^{n}m^2(x_i,\theta)W_{ji}^2\right]^2 \\
&= \sum_{j=1}^{n-1}m^4(x_j,\theta)\sum_{i=j+1}^{n}m^4(x_i,\theta)W_{ji}^4 + 2\sum_{j=1}^{n-2}m^4(x_j,\theta)\sum_{i=j+1}^{n-1}\sum_{k>i}m^2(x_i,\theta)m^2(x_k,\theta)W_{ji}^2W_{jk}^2 \\
&\equiv V_{n1}^{(1)} + V_{n1}^{(2)} \tag{41}
\end{aligned}
$$

Next, in (42) the first equality follows from direct calculation, while the second is implied by the change of variables $u = (z_i - z_j)/h$. The final equality is then implied by Assumption 2(ii) which permits the use of the dominated convergence theorem.

$$
\begin{aligned}
\frac{n(n-1)}{h^{d_z}}E[V_{n1}^{(1)}] &= \frac{h^{-d_z}}{2}\int E[m^4(X_i,\theta)|z_i]E[m^4(X_j,\theta)|z_j]K^4\left(\frac{z_i-z_j}{h}\right)f_Z(z_i)f_Z(z_j)dz_idz_j \\
&= \frac{1}{2}\int E[m^4(X_i,\theta)|z_i]E[m^4(X_i,\theta)|z_i-hu]K^4(u)f_Z(z_i-hu)f_Z(z_j)dudz_i \\
&= \frac{1}{2}E\left[\left(E[m^4(X_i,\theta)|Z_i]\right)^2 f_Z(Z)\right]\left[\int K^4(u)du\right] + o(1) \tag{42}
\end{aligned}
$$

Therefore, arguing as in (36), (37) and (39) we conclude (43).

$$n^{-4}h^{-2d_z}V_{n1}^{(1)} \xrightarrow{a.s.} 0 \tag{43}$$

To examine $V_{n1}^{(2)}$, define the non-symmetric kernel

$$\tilde{H}_n^1(w_j, w_i, w_k) = m^4(x_j,\theta)m^2(x_i,\theta)m^2(x_k,\theta)W_{ji}^2W_{jk}^2 \tag{44}$$

Hence, by definition the first equality in (45) is implied. Next, notice that since every term in the summation is positive, $V_{n1}^{(2)}$ is smaller than the symmetric U-Statistic obtained in the first inequality in (45). We denote the resulting expression as $\tilde{V}_{n1}^{(2)}$.

$$
\begin{aligned}
V_{n1}^{(2)} &= 2\sum_{j=1}^{n-2}\sum_{i=j+1}^{n-1}\sum_{k>i}\tilde{H}_n^1(w_j, w_i, w_k) \\
&\leq 2\sum_{j=1}^{n-2}\sum_{i=j+1}^{n-1}\sum_{k>i}\tilde{H}_n^1(w_j, w_i, w_k) + \tilde{H}_n^1(w_i, w_j, w_k) + \tilde{H}_n^1(w_k, w_j, w_i) \\
&\equiv \tilde{V}_{n1}^{(2)} \tag{45}
\end{aligned}
$$

Let $P_n^i(\tilde{V}_1^{(2)})$ denote the ith term in the Hoeffding decomposition of $\tilde{V}_{n1}^{(2)}$. Through calculations analogous to (33) and (38), which we omit for the sake of brevity, it is possible to establish (46).

$$
\begin{array}{ll}
n^{-4}h^{-2d_z}P_n^0(\tilde{V}_1^{(2)}) = O(n^{-1}) & n^{-8}h^{-4d_z}E[(P_n^1(\tilde{V}_1^{(2)}))^2] = O(n^{-3}) \\
n^{-8}h^{-4d_z}E[(P_n^2(\tilde{V}_1^{(2)}))^2] = O(n^{-4}h^{-d_z}) & n^{-8}h^{-4d_z}E[(P_n^3(\tilde{V}_1^{(2)}))^2] = O(n^{-5}h^{-2d_z})
\end{array} \tag{46}
$$

16

Hence, arguing as in (36), (37) and (39) we conclude $n^{-4}h^{-2d_z}\tilde{V}_{n1}^{(2)} \xrightarrow{a.s.} 0$. Therefore, together with (45), (43) and (41) this result implies (47).

$$n^{-4}h^{-2d_z}E^*\left[(V_{n1}^* - E^*[V_{n1}^*])^2\right] \xrightarrow{a.s.} 0 \tag{47}$$

To conclude, note that (40) and (47) imply $n^{-2}h^{-d_z}V_{n1} \xrightarrow{p^*} 0$ $a.s.$ as desired.

We now establish the second claim of the Lemma. First note that since the $m^*(x_i,\theta)$ i.i.d. and $E^*[m^*(x_i,\theta)] = 0$ result (48) is implied.

$$E^*[V_{n2}^*] = 0 \tag{48}$$

In (49), we exchange the order of summation in order to obtain the first equality. In turn, the second equality is implied by $E^*[(m^*(x_i,\theta))^2] = m^2(x_i,\theta)$, while the third equality follows from direct calculation. We denote the terms in the resulting decomposition as $V_{n2}^{(1)}$ and $V_{n2}^{(2)}$.

$$
\begin{aligned}
E^*\left[(V_{n2}^*)^2\right] &= 4E^*\left[\left(\sum_{j=2}^{n-1}\sum_{k<j}m^*(x_j,\theta)m^*(x_k,\theta)\sum_{i=j+1}^{n}m^2(x_i,\theta)W_{ij}W_{ik}\right)^2\right] \\
&= 4\sum_{j=2}^{n-1}\sum_{k<j}m^2(x_j,\theta)m^2(x_k,\theta)\left[\sum_{i=j+1}^{n}m^2(x_i,\theta)W_{ij}W_{ik}\right]^2 \\
&= 4\sum_{j=2}^{n-1}\sum_{k<j}m^2(x_j,\theta)m^2(x_k,\theta)\left[\sum_{i=j+1}^{n}m^4(x_i,\theta)W_{ij}^2W_{ik}^2 + 2\sum_{i=j+1}^{n-1}\sum_{l>i}m^2(x_i,\theta)m^2(x_l,\theta)W_{ij}W_{ik}W_{lj}W_{lk}\right] \\
&\equiv V_{n2}^{(1)} + V_{n2}^{(2)} \tag{49}
\end{aligned}
$$

Notice, that $V_{n2}^{(1)} = 2V_{n1}^{(2)}$, as defined in (41). As already shown, however, $n^{-4}h^{-2d_z}V_{n1}^{(2)} \xrightarrow{a.s.} 0$, and therefore we conclude (50).

$$n^{-4}h^{-2d_z}V_{n2}^{(1)} \xrightarrow{a.s.} 0 \tag{50}$$

To examine, $V_{n2}^{(2)}$, define the non-symmetric kernel

$$\tilde{H}_n^2(w_j,w_k,w_i,w_l) = m^2(x_j,\theta)m^2(x_k,\theta)m^2(x_i,\theta)m^2(x_l,\theta)|W_{ij}W_{ik}W_{lj}W_{lk}| \tag{51}$$

Thus, by definition, the first inequality in (52) follows. The second inequality in (52) then follows by $\tilde{H}_n^2(w_j,w_k,w_i,w_l) \geq 0$. To conclude, notice that the resulting U-Statistic is symmetric in its arguments, and denote it $\tilde{V}_{n2}^{(2)}$.

$$
\begin{aligned}
|V_{n2}^{(2)}| &\leq 8\sum_{j=2}^{n-2}\sum_{k<j}\sum_{i=j+1}^{n-1}\sum_{l>i}\tilde{H}_n^2(w_j,w_k,w_i,w_l) \\
&\leq 8\sum_{j=2}^{n-2}\sum_{k<j}\sum_{i=j+1}^{n-1}\sum_{l>i}\tilde{H}_n^2(w_j,w_k,w_i,w_l) + \tilde{H}_n^2(w_i,w_k,w_j,w_l) + \tilde{H}_n^2(w_l,w_k,w_i,w_j) \\
&\equiv \tilde{V}_{n2}^{(2)} \tag{52}
\end{aligned}
$$

Let $P_n^i(\tilde{V}_2^{(2)})$ denote the ith term in the Hoeffding decomposition of $\tilde{V}_{n2}^{(2)}$. Through calculations analogous to (33) and (38), which we omit for the sake of brevity, it is possible to establish (53).

$$
\begin{aligned}
&n^{-4}h^{-2d_z}P_n^0(\tilde{V}_2^{(2)}) = O(h^{d_z}) && n^{-8}h^{-4d_z}E[(P_n^1(\tilde{V}_2^{(2)}))^2] = O(n^{-1}h^{2d_z}) \\
&n^{-8}h^{-4d_z}E[(P_n^2(\tilde{V}_2^{(2)}))^2] = O(n^{-2}h^{d_z}) && n^{-8}h^{-4d_z}E[(P_n^3(\tilde{V}_2^{(2)}))^2] = O(n^{-3}) \\
&n^{-8}h^{-4d_z}E[(P_n^4(\tilde{V}_2^{(2)}))^2] = O(n^{-4}h^{-d_z})
\end{aligned}
\tag{53}
$$

Therefore, arguing as in (36), (37) and (39) we conclude $n^{-4}h^{-2d_z}\tilde{V}_{n2}^{(2)} \xrightarrow{a.s.} 0$. Together with (52), (50) and (49), this implies (54).

$$n^{-4}h^{-2d_z}E^*[(V_{n2}^*)^2] \xrightarrow{a.s.} 0 \tag{54}$$

To conclude, note that (48) and (54) imply $n^{-2}h^{-d_z}V_{n2}^* \xrightarrow{p^*} 0$ a.s., which concludes the proof of the Lemma. ∎

Lemma .2. *Let $\mathcal{F}_i$ denote the $\sigma-$field generated by $\{m^*(X_1,\theta),\ldots,m^*(X_i,\theta)\}$ conditional on the sample. If Assumptions 1(i)-(iii), 2(ii)-(iii) and 3(i)-(ii) hold with $\bar{u} \geq 10$, then for any $\epsilon > 0$, Y_i^* as defined in (22) satisfies*

$$n^{-2}h^{-d_z}\sum_{i=2}^{n} E^* \left[(Y_i^*)^2 1\{|Y_i^*| > \epsilon n h^{\frac{d_z}{2}}\}|\mathcal{F}_{i-1}\right] \xrightarrow{p^*} 0 \quad a.s.$$

Proof of Lemma .2: Let $S_i^* = \sum_{j=1}^{i-1} W_{ij}m^*(x_j,\theta)$, and note that $Y_i^* = m^*(x_i,\theta)S_i^*$. Therefore, for any $k > 0$, the event $|Y_i^*| > \epsilon n h^{\frac{d_z}{2}}$ implies either $|m^*(x_i,\theta)| > k$ or $k|S_i^*| > \epsilon n h^{\frac{d_z}{2}}$. Use $E^*[(m^*(x_i,\theta))^2] = m^2(x_i,\theta)$ and S_i^* being independent of $m^*(x_i,\theta)$ to derive the first inequality in (55). For the second inequality, apply the Cauchy-Schwarz and Chebychev's inequalities and notice that $E^*[(m^*(x_i,\theta))^4] = 2m^4(x_i,\theta)$ and $E^*[|m^*(x_i,\theta)|] \leq |m(x_i,\theta)|$.

$$\frac{1}{n^2h^{d_z}}\sum_{i=2}^{n}E^* \left[(Y_i^*)^2 1\{|Y_i^*| > \epsilon n h^{\frac{d_z}{2}}\}\right]$$
$$\leq \frac{1}{n^2h^{d_z}}\sum_{i=2}^{n} E^* \left[(S_i^*)^2(m^*(x_i,\theta))^2 1\{|m^*(x_i,\theta)| > k\}\right] + \frac{1}{n^2h^{d_z}}\sum_{i=2}^{n} m^2(x_i,\theta)E^* \left[(S_i^*)^2 1\{k|S_i^*| > \epsilon n h^{\frac{d_z}{2}}\}\right]$$
$$\leq \frac{2}{n^2h^{d_z}\sqrt{k}}\sum_{i=2}^{n} |m(x_i,\theta)|^{\frac{5}{2}}E^*[(S_i^*)^2] + \frac{k^2}{\epsilon^2 n^4 h^{2d_z}}\sum_{i=2}^{n} m^2(x_i,\theta)E^* \left[(S_i^*)^4\right] \tag{55}$$

We will establish the Lemma by studying the two resulting terms in (55). First note that since $E^*[m^*(x_i,\theta)] = 0$ and $E^*[(m^*(x_i,\theta))^2] = m^2(x_i,\theta)$, after expanding the square the first equality in (56) immediately follows. We denote the resulting statistic by L_{n1}.

$$\frac{2}{n^2h^{d_z}\sqrt{k}}\sum_{i=2}^{n} |m(x_i,\theta)|^{\frac{5}{2}}E^*[(S_i^*)^2] = \frac{2}{n^2h^{d_z}\sqrt{k}}\sum_{i=2}^{n}\sum_{j=1}^{i-1} |m(x_i,\theta)|^{\frac{5}{2}}m^2(x_j,\theta)W_{ij}^2 \equiv L_{n1} \tag{56}$$

Similarly, expand the fourth power to obtain the equality in (57) and denote the resulting statistics by $L_{n2}^{(1)}$ and $L_{n2}^{(2)}$.

$$\frac{k^2}{\epsilon^2 n^4 h^{2d_z}}\sum_{i=2}^{n} m^2(x_i,\theta)E^* \left[(S_i^*)^4\right]$$
$$= \frac{2k^2}{\epsilon^2 n^4 h^{2d_z}}\sum_{i=2}^{n} m^2(x_i,\theta)\sum_{j=1}^{i-1} m^4(x_j,\theta)W_{ij}^4 + \frac{2k^2}{\epsilon^2 n^4 h^{2d_z}}\sum_{i=3}^{n} m^2(x_i,\theta)\sum_{j=2}^{i-1}\sum_{k<j} m^2(x_j,\theta)m^2(x_k,\theta)W_{ij}^2 W_{ik}^2$$
$$\equiv L_{n2}^{(1)} + L_{n2}^{(2)} \tag{57}$$

Next, proceed as in (45) to see L_{n1} is dominated by a symmetric U-Statistic. Then argue as in (29), (33), (37) and (39) to establish (58).

$$L_{n1} \xrightarrow{a.s.} \frac{1}{\sqrt{k}}E \left[E[m^2(X,\theta)|Z]E[|m(X_i,\theta)|^{\frac{5}{2}}|Z]f_Z(Z)\right] \left[\int K^2(u)du\right] \tag{58}$$

Next, notice the similarity of $L_{n2}^{(1)}$ and $L_{n2}^{(2)}$ to $V_{n1}^{(1)}$ and $V_{n1}^{(2)}$ as defined in (41). By the same arguments that lead to (47), it is straightforward to establish (59).

$$L_{n2}^{(1)} + L_{n2}^{(2)} \xrightarrow{a.s.} 0 \tag{59}$$

18

Therefore, since k can be chosen arbitrarily, combining (55), (58) and (59) establishes (60).

$$\frac{1}{n^2 h^{d_z}} \sum_{i=2}^{n} E^* \left[(Y_i^*)^2 1\{|Y_i^*| > \epsilon n h^{\frac{d_z}{2}}\} \right] \xrightarrow{a.s.} 0 \tag{60}$$

To conclude, note that (60) and Markov's inequality imply the desired result. $\blacksquare$

APPENDIX C - Proof of Lemma 3.2, Auxiliary Theorems .1, .2 and Auxiliary Lemma .3

Proof of Lemma 3.2: We omit the proof of a multidimensional analogue to Lemma 3.1, which implies convergence of the finite dimensional distributions of the bootstrap. Also note that as implied by Lemma 2.1 and Lemma 3.1, the Gaussian processes $G(\theta)$ and $G^b(\theta)$ have the same marginals on $\mathcal{L}^\infty(\Theta_0)$. Hence, due to Theorem 1.5.4 and 1.5.7 in van der Vaart & Wellener (1997), it only remains to show $a.s.$ uniform asymptotic equicontinuity in probability of $T_n^*(\theta)$ with respect to the norm $||\cdot||_\infty$. That is, we wish to show that for every $\epsilon, \eta > 0$, there exists a $\delta > 0$ such that (61) holds.

$$\limsup_{n \to \infty} P^* \left(\sup_{||\theta_1 - \theta_2||_\infty < \delta} |T_n^*(\theta_1) - T_n^*(\theta_2)| > \epsilon \right) < \eta \quad a.s. \tag{61}$$

To establish this result, we first derive a maximal inequality for the process $T_n^*(\theta)$. Define,

$$a_{ij}(\theta_1, \theta_2) = h^{-\frac{d_z}{2}} K\left(\frac{z_i - z_j}{h}\right) (m(x_i, \theta_1)m(x_j, \theta_1) - m(x_i, \theta_2)m(x_j, \theta_2)) \tag{62}$$

The first equality in (63) then follows by definition. Next note that the U-Statistics are P^*-canonical and apply the decoupling inequalities from de la Pena (1992), also Proposition 2.1 in Arcones & Gine (1993), to obtain the inequality in (63) for $\{\epsilon_i\}_{i=1}^n$ i.i.d. rademacher random variables.

$$E^* \left[\sup_{||\theta_1 - \theta_2||_\infty < \delta} |T_n^*(\theta_1) - T_n^*(\theta_2)| \right] = E^* \left[\sup_{||\theta_1 - \theta_2||_\infty < \delta} \left| \frac{2}{n-1} \sum_{i=2}^{n} \sum_{j<i} a_{ij}(\theta_1, \theta_2) u_i u_j \right| \right]$$

$$\lesssim E^* \left[\sup_{||\theta_1 - \theta_2||_\infty < \delta} \left| \frac{1}{n-1} \sum_{i=2}^{n} \sum_{j<i} a_{ij}(\theta_1, \theta_2) u_i u_j \epsilon_i \epsilon_j \right| \right] \tag{63}$$

Define the random semimetrics:

$$(\tilde{\rho}_{2n}^*(\theta_1, \theta_2))^2 = \frac{2}{n(n-1)} \sum_{i=2}^{n} \sum_{j<i} a_{ij}^2(\theta_1, \theta_2) u_i^2 u_j^2 \qquad (\tilde{\rho}_{2n}(\theta_1, \theta_2))^2 = \frac{2}{n(n-1)} \sum_{i=2}^{n} \sum_{j<i} a_{ij}^2(\theta_1, \theta_2) \tag{64}$$

Let $E_\epsilon[\cdot]$ denote the expectation over the rademacher random variables. In addition, define the class of functions $\Theta_\delta = \{\theta_1(x) - \theta_2(x) : ||\theta_1 - \theta_2||_\infty < \delta\}$. The inequality in (65) then follows by applying Propositions 2.2 and 2.6 in Arcones & Gine (1993), where D_n^* is the diameter of Θ_δ under $\tilde{\rho}_{2n}^*(\theta_1, \theta_2)$.

$$E_\epsilon \left[\sup_{||\theta_1 - \theta_2||_\infty < \delta} \left| \frac{1}{n-1} \sum_{i=2}^{n} \sum_{j<i} a_{ij}(\theta_1, \theta_2) u_i u_j \epsilon_i \epsilon_j \right| \right] \lesssim \int_0^{D_n^*} \log N(\Theta_\delta, \tilde{\rho}_{2n}^*, \epsilon) d\epsilon \tag{65}$$

Combining (63) and (65) yields the first inequality in (66). Furthermore, since $U^2 \leq 3$, the definitions in (64) imply $\tilde{\rho}^*(\theta_1, \theta_2) \leq 3\tilde{\rho}(\theta_1, \theta_2)$. Hence, the second inequality in (66) follows for D_n the diameter of Θ_δ under $\tilde{\rho}_{2n}(\theta_1, \theta_2)$.

$$E^* \left[\sup_{||\theta_1 - \theta_2||_\infty < \delta} |T_n^*(\theta_1) - T_n^*(\theta_2)| \right] \lesssim E^* \left[\int_0^{D_n^*} \log N(\Theta_\delta, \tilde{\rho}_{2n}^*, \epsilon) d\epsilon \right]$$

$$\leq \int_0^{3D_n} \log N(\Theta_\delta, 3\tilde{\rho}_{2n}, \epsilon) d\epsilon \tag{66}$$

19

In addition, notice that Lemma .3 implies the inequality in (67) for any $\theta_1, \theta_2 \in \Theta_\delta$.

$$\tilde{\rho}_{2n}(\theta_1, \theta_2) \leq h^{\frac{d_z}{2}} ||J_n||_{2n} ||\theta_1 - \theta_2||_\infty \tag{67}$$

Furthermore, if $\theta_1 \in \Theta_\delta$, then it is of the form $\theta_1(x) = \theta_{11}(x) - \theta_{12}(x)$ with $||\theta_{11} - \theta_{12}||_\infty < \delta$. Hence, since $D_n = \sup_{\theta_1, \theta_2 \in \Theta_\delta} \tilde{\rho}(\theta_1, \theta_2)$, (67) implies $D_n \leq 2\delta h^{\frac{d_z}{2}} ||J_n||_{2n}$. Thus, the first inequality in (68) follows by (67). For the second inequality in (68) observe that for any norm $|| \cdot ||$, it follows that $N(\Theta_\delta, || \cdot ||, \epsilon) \leq N^2(\Theta, || \cdot ||, \epsilon/2)$. In turn the third inequality is implied by the change of variables $u = \epsilon/6h^{\frac{d_z}{2}} ||J_n||_{2n}$.

$$\int_0^{3D_n} \log N(\Theta_\delta, 3\tilde{\rho}_{2n}, \epsilon) d\epsilon \leq \int_0^{6\delta h^{\frac{d_z}{2}} ||J_n||_{2n}} \log N(\Theta_\delta, || \cdot ||_\infty, \epsilon/3h^{\frac{d_z}{2}} ||J_n||_{2n}) d\epsilon$$
$$\leq 2 \int_0^{6\delta h^{\frac{d_z}{2}} ||J_n||_{2n}} \log N(\Theta, || \cdot ||_\infty, \epsilon/6h^{\frac{d_z}{2}} ||J_n||_{2n}) d\epsilon$$
$$\leq 6h^{\frac{d_z}{2}} ||J_n||_{2n} \int_0^\delta \log N(\Theta, || \cdot ||_\infty, u) du \tag{68}$$

Let $P_n^i(J^2)$ denote the ith term in the Hoeffding decomposition of the U-Statistic $||J_n||_{2n}^2 h^{d_z}$. The first equality in (69) then follows by direct calculation. Furthermore, the arguments in (80) imply that if $0 \leq j \leq \bar{u}/2$, then $E[(E[J_n^2|W_i])^j] h^{jd_z} = O(1)$. Together with Lemma .3 this implies the second equality in (69).

$$E\left[(P_n^1(J^2))^4\right] = \frac{h^{4d_z}}{n^3} E\left[(E[J_n^2|W_i] - E[J_n^2])^4\right] + \frac{h^{4d_z}(n-1)}{2n^3}\left(E\left[(E[J_n^2|W_i] - E[J_n^2])^2\right]\right)^2$$
$$= O(n^3) + O(n^2) \tag{69}$$

Therefore, utilizing the Borel-Cantelli Lemma as in (36) and (37) and employing result (69) establishes (70).

$$P_n^1(J^2) \xrightarrow{a.s.} 0 \tag{70}$$

We now examine $P_n^2(J^2)$. The first equality and inequality in (71) follow by direct calculation. In turn, the second equality in (71) follows by Lemma .3 and the same arguments as in (69).

$$E\left[(P_n^2(J^2))^2\right] = \frac{2}{n(n-1)} h^{2d_z} E\left[(J_n^2(W_i, W_j) - E[J_n^2|W_i] - E[J_n^2|W_j] + E[J_n^2])^2\right]$$
$$\leq \frac{32h^{2d_z}}{n(n-1)}\left[E[J_n^4(W_i, W_j)] + 2E\left[(E[J_n^2|W_i])^2\right] + (E[J_n^2])^2\right]$$
$$= O(h^{d_z} n^2) + O(n^2) + O(n^2) \tag{71}$$

Next notice that Assumptions 3(ii)-(iii) imply $n^\delta = o(nh^{d_z})$ for some $\delta > 0$. Hence, by (69) $E\left[(P_n^2(J^2))^2\right] = O(n^{1+\delta})$ for some $\delta > 0$. Applying the Borel Cantelli Lemma as in as in (36) and (37) therefore establishes (72).

$$P_n^2(J^2) \xrightarrow{a.s.} 0 \tag{72}$$

Using the Hoeffding decomposition of $||J_n^2||_{2n}^2 h^{d_z}$ in turn implies the first equality in (73). The inequality in (73) the holds almost surely for n sufficiently large as a result of (70), (71) and (81).

$$h^{d_z} ||J_n||_{2n}^2 = h^{d_z} E[J_n^2] + \sum_{i=1}^2 P_n^i(J^2) \leq 2\left[\int K^2(u) du\right] \left[E[F^2(X)E[G^2(X)|Z]f_Z(Z)]\right] \quad a.s. \tag{73}$$

To conclude, use Markov's inequality, (66) and (68) to establish the first inequality in (74). The second inequality then holds almost surely for some $M > 0$ due to result (73). In turn, the first equality is implied by Theorem .4 in

Santos (2007), while the final result is due to $(m_0 + \delta_0)d_x/(m_0\delta_0) < 1$.

$$\lim_{\delta\downarrow 0}\lim\sup_{n\to\infty} P^*\left(\sup_{||\theta_1-\theta_2||_\infty<\delta}|T_n^*(\theta_1) - T_n^*(\theta_2)| > \epsilon\right) \lesssim \lim_{\delta\downarrow 0}\lim\sup_{n\to\infty}\frac{h^{\frac{d_z}{2}}}{\epsilon}||J_n||_{2n}\int_0^\delta \log N(\Theta, ||\cdot||_\infty, u)du$$

$$\leq \lim_{\delta\downarrow 0} M\int_0^\delta \log N(\Theta, ||\cdot||_\infty, u)du \quad a.s.$$

$$= \lim_{\delta\downarrow 0} M\int_0^\delta u^{-\frac{(m_0+\delta_0)d_x}{m_0\delta_0}}du$$

$$= 0 \tag{74}$$

Result (74) implies $T_n^*(\theta)$ is $a.s.$ uniformly asymptotic equicontinuous which establishes the Lemma. ■

Theorem .1. *Let $\mathcal{F}$ be a set of P canonical symmetric functions with envelope F and $||F||_{\mathcal{L}^P(\mathcal{X}^m)} < \infty$. Define I_m^n to be the set of distinct m-tuples from a sample of size n, $U_n^m(f) = n^{-m}\sum_{I_m^n} f(x_{i_1}, \ldots, x_{i_m})$ and D_n to be the diameter of $\mathcal{F}$ under ρ_{mn}. Then there exists a constant K depending only on m and p such that:*

$$E\left[\left(\sup_{f,g\in\mathcal{F}} n^{\frac{m}{2}}|U_n^m(f) - U_n^m(g)|\right)^p\right] \leq KE\left[\left(\int_0^{D_n}[\log N(\mathcal{F}, \rho_{mn}, \epsilon)]^{\frac{m}{2}}d\epsilon\right)^p\right]$$

Proof of Theorem .1: This result follows arguments similar to Theorem 3.2 in Arcones & Gine (1994). The constants C, C' may change from line to line throughout the proof, but they always denote constants dependent only on m and p. The case $p = 1$ is shown in Arcones & Gine (1993), so we assume $p > 1$. First we use a results from de la Pena (1992), Proposition 2.1 in Arcones & Gine (1993), to derive that for $\mathcal{F}$ a class of P-canonical, symmetric functions with envelope F satisfying $||F||_{\mathcal{L}^P(\mathcal{X}^m)} < \infty$ we have:

$$E\left[\sup_{f,g\in\mathcal{F}}\left(n^{\frac{m}{2}}|U_n^m(f) - U_n^m(g)|\right)^p\right] = E\left[\sup_{f,g\in\mathcal{F}}\left(n^{-\frac{m}{2}}\left|\sum_{I_m^n} f(x_{i_1}, \ldots, x_{i_m}) - g(x_{i_1}, \ldots, x_{i_m})\right|\right)^p\right]$$

$$\leq CE\left[\sup_{f,g\in\mathcal{F}}\left(n^{-\frac{m}{2}}\left|\sum_{I_m^n}\epsilon_{i_1}\ldots\epsilon_{i_m}(f(x_{i_1}, \ldots, x_{i_m}) - g(x_{i_1}, \ldots, x_{i_m}))\right|\right)^p\right] \tag{75}$$

where $\{\epsilon_i\}_{i=1}^n$ are i.i.d. Rademacher random variables independent of $\{X_i\}_{i=1}^n$. Condition on $\{X_i\}_{i=1}^n$ and let $E_\epsilon[\cdot]$ denote the expectation over $\{\epsilon_i\}_{i=1}^n$. Let $\tilde{U}_n^m(f) = n^{-\frac{m}{2}}\sum_{I_m^n}\epsilon_{i_1}\ldots\epsilon_{i_m}f(x_{i_1}, \ldots, x_{i_m})$. By Proposition 2.2 in Arcones & Gine (1993), originally in Borell (1979), we have for any $1 < q < p < \infty$:

$$\left(E_\epsilon[|\tilde{U}_n^m(f)|^p]\right)^{\frac{1}{p}} \leq \left(\frac{p-1}{q-1}\right)^{\frac{m}{2}}\left(E_\epsilon[|\tilde{U}_n^m(f)|^q]\right)^{\frac{1}{q}} \tag{76}$$

Hence, using Holder's inequality and (76) for the first and second inequalities in (77) respectively, we conclude:

$$E_\epsilon[|\tilde{U}_n^m(f)|^p] = E_\epsilon\left[|\tilde{U}_n^m(f)|^{\frac{1}{2}}|\tilde{U}_n^m(f)|^{\frac{2p-1}{2}}\right] \leq \left(E_\epsilon\left[|\tilde{U}_n^m(f)|\right]\right)^{\frac{1}{2}}\left(E_\epsilon\left[|\tilde{U}_n^m(f)|^{2p-1}\right]\right)^{\frac{1}{2}}$$

$$\leq C\left(E_\epsilon\left[|\tilde{U}_n^m(f)|\right]\right)^{\frac{1}{2}}\left(E_\epsilon\left[|\tilde{U}_n^m(f)|^p\right]\right)^{\frac{2p-1}{2p}} \tag{77}$$

Applying (77), we obtain the first inequality in (78). The second inequality is derived in Theorem .2 in Santos (2007) following Arcones & Gine (1993).

$$E_\epsilon\left[\sup_{f,g\in\mathcal{F}}\left|\tilde{U}_n^m(f-g)\right|^p\right] \leq C\left(E_\epsilon\left[\sup_{f,g\in\mathcal{F}}\left|\tilde{U}_n^m(f-g)\right|\right]\right)^p \leq C'\left(\int_0^{D_n}[\log N(\mathcal{F}, \rho_{mn}(f,g), \epsilon)]^{\frac{m}{2}}d\epsilon\right)^p \tag{78}$$

To conclude the proof, take the expectation of (78) over $\{X_i\}_{i=1}^n$ and use (75). ■

Theorem .2. *Let $\mathcal{F} = \{f_t : t \in T\}$ be a class of functions such that $|f_s(x) - f_t(x)| \leq d(s,t)F(x)$ for every s and t and some fixed function F. Then for any norm $||\cdot||$,*

$$N_{[\,]}(\mathcal{F}, ||\cdot||, 2\epsilon||F||) \leq N(T, d, \epsilon)$$

Proof of Theorem .2: Refer to Theorem 2.7.11 in van der Vaart & Wellner. ∎

Lemma .3. *If Assumptions 2(ii)-(iii) and 3(i)-(ii) hold, then there are $J_n(w_i, w_j)$ such that*

$$|H_n(w_i, w_j, \theta_1) - H_n(w_i, w_j, \theta_2)| \leq J_n(w_i, w_j)||\theta_1 - \theta_2||_\infty$$

In addition, the expectations $E[(E[J_n|W_i])^j]$ and $h^{d_z(j-1)}E[J_n^j]$ are all uniformly bounded in n for $1 \leq j \leq \bar{u}$.

Proof of Lemma .3: For the first claim of the Lemma, we use Assumptions 2(ii)-(iii) to obtain the first inequality in (79) and define the resulting function to be $J_n(w_i, w_j)$.

$$|H_n(w_i, w_j, \theta_1) - H_n(w_i, w_j, \theta_2)| \leq h^{-d_z}\left|K\left(\frac{z_i - z_j}{h}\right)\right|(F(x_i)G(x_j) + F(x_j)G(x_i))\,||\theta_1 - \theta_2||_\infty$$

$$\equiv J_n(w_i, w_j)||\theta_1 - \theta_2||_\infty \tag{79}$$

In (80) we verify $E[(E[J_n|W_i])^j]$ is uniformly bounded in n for $1 \leq j \leq \bar{u}$. The first inequality follows by convexity, since $j \geq 1$, while the second inequality is implied by $E[G(X_j)|z_j]$ and $E[F(X_j)|z_j]$ being bounded. For the first equality do the change of variables $u = (z_i - z_j)/h$, and for the second equality notice we can apply the dominated convergence theorem due to Assumptions 2(ii)-(iii).

$$E[(E[J_n|W_i])^j] \leq 2^j E\left[\left(h^{-d_z}E\left[\left|K\left(\frac{Z_i - Z_j}{h}\right)\right|F(X_i)G(X_j)|W_i\right]\right)^j + \left(h^{-d_z}E\left[\left|K\left(\frac{Z_i - Z_j}{h}\right)\right|F(X_j)G(X_i)|W_i\right]\right)^j\right]$$

$$\lesssim E\left[\left(h^{-d_z}E\left[\left|K\left(\frac{Z_i - Z_j}{h}\right)\right||W_i\right]\right)^j(F^j(X_i) + G^j(X_i))\right]$$

$$= \int E[F^j(X_i)|z_i]E[G^j(X_i)|z_i]\left(\int|K(u)|f_Z(z_i - hu)du\right)^j f_Z(z_i)dz_i$$

$$= \left[\int|K(u)|du\right]E[F^j(X)E[G^j(X)|Z]f_Z^j(Z)] + o(1) \tag{80}$$

For the second claim of the Lemma, we examine $h^{d_z(j-1)}E[J_n^j]$ in (81). The first inequality follows by convexity since $j \geq 1$, while the first equality is implied by the change of variables $u = (z_i - z_j)/h$. For the final equality use that $E[G^j(X)|z]$ and $f_Z(z)$ are bounded to apply the dominated convergence theorem.

$$h^{d_z(j-1)}E[J_n^j] \leq h^{-d_z}2^j E\left[\left|K^j\left(\frac{Z_i - Z_j}{h}\right)\right|F^j(X_i)G^j(X_j)\right]$$

$$= \int|K^j(u)|E[F^j(X_i)|z_i]E[G^j(X_i)|z_i - hu]f_Z(z_i - hu)f_Z(z_i)dudz_i$$

$$= \left[\int|K^j(u)|du\right]\left[E[F^j(X)E[G^j(X)|Z]f_Z(Z)]\right] + o(1) \tag{81}$$

The resulting expectations in (80) and (81) are finite by Assumptions 2(ii)-(iii), which establishes the Lemma. ∎

APPENDIX D - Proof of Theorems 3.1, 3.2, Auxiliary Theorem .3 and Auxiliary Lemma .4

Theorem .3. *If Assumptions 1(i)-(iv), 2(i)-(iii), 3(i)-(iii), 4(i) hold and $\Theta_0 \cap R \neq \emptyset$, then*

$$\inf_{\Theta \cap R} \left(n^{-r} T_n(\theta) + T_n^*(\theta) \right) \xrightarrow{\mathcal{L}^*} \inf_{\Theta_0 \cap R} G(\theta) \quad a.s.$$

where $G(\theta)$ is a Gaussian process on $\mathcal{L}^\infty(\Theta_0)$ with the same marginals as in Theorem 2.1. Furthermore, under the same assumptions, if $\Theta_0 \cap R = \emptyset$, then

$$\inf_{\Theta \cap R} \left(n^{-r} T_n(\theta) + T_n^*(\theta) \right) \xrightarrow{p^*} \infty \quad a.s.$$

Proof of Theorem .3: The strategy of the proof is to show that since the term $n^{-r} T_n(\theta)$ diverges to infinity outside a shrinking neighborhood of Θ_0, the minimum is asymptotically attained in Θ_0. Because in $\mathcal{L}^\infty(\Theta_0)$, the law of $T_n^*(\theta)$ is consistent for that of $G(\theta)$ in Theorem 2.1, the result will follow. With this purpose define:

$$\Theta_0^{\epsilon_n} = \{ \theta \in \Theta : E[(E[m(X,\theta)|Z])^2 f_Z(Z)] \leq \epsilon_n \} \tag{82}$$

where $\epsilon_n \searrow 0$ with ϵ_n proportional to h^l for $l \leq \lfloor k \rfloor$, $n^{\frac{1}{2}-r} h^l \to \infty$ and $n^{\gamma+(1-2r)\frac{\bar{u}}{2}} h^{\left(d_z + l - \frac{l(m_0+\delta_0)d_x}{2m_0\delta_0} \right) \frac{\bar{u}}{2}} \to 0$ for some $\gamma > 1$. Notice this is possible due to Assumption 4(i).

First assume $\Theta_0 \cap R \neq \emptyset$. In this case, we begin by deriving a useful inequality in (83). The left hand side holds by simple manipulations, while the right hand side is implied by $\Theta_0^{\epsilon_n} \subseteq \Theta_0$.

$$-\sup_{\Theta_0^{\epsilon_n}} n^{-r} |T_n(\theta) - E[T_n(\theta)]| + \inf_{\Theta_0^{\epsilon_n}} n^{-r} E[T_n(\theta)] + \inf_{\Theta_0^{\epsilon_n} \cap R} T_n^*(\theta)$$

$$\leq \inf_{\Theta_0^{\epsilon_n} \cap R} \left(n^{-r} T_n(\theta) + T_n^*(\theta) \right) \leq \sup_{\Theta_0 \cap R} n^{-r} T_n(\theta) + \inf_{\Theta_0 \cap R} T_n^*(\theta) \tag{83}$$

In (84), we examine $E[H_n(\theta)]$. The first equality follows by the change of variables $u = (z_i - z_j)/h$, while the second equality follows by a $\lfloor k \rfloor$ order Taylor expansion of the integral around $h = 0$ and the fact that $K(u)$ is a kernel of order $\lfloor k \rfloor$. Differentiation through the integral is permitted due to Assumptions 2(i) and 3(i).

$$\left| E[H_n(\theta)] - E\left[(E[m(X,\theta)|Z])^2 f_Z(Z) \right] \right|$$

$$= \left| \int K(u) E[m(X_i,\theta)|z_i - hu] E[m(X_i,\theta)|z_i] f_Z(z_i) f_Z(z_i - hu) dz_i du - E\left[(E[m(X,\theta)|Z])^2 f_Z(Z) \right] \right|$$

$$= h^{\lfloor k \rfloor} \times \left| \int K(u) E[m(X,\theta)|z_i] (\nabla_h^{\lfloor k \rfloor} E[m(X,\theta)|z_i - \tilde{h}u] f_Z(z_i - \tilde{h}u)) f_Z(z_i) dz_i du \right| \tag{84}$$

Next, notice that since $E[T_n(\theta)] = 0$ for $\theta \in \Theta_0$, the first inequality in (85) immediately follows. The second inequality in (85) is then implied by $E[T_n(\theta)] = h^{\frac{d_z}{2}} n E[H_n(\theta)]$. Use (84) to obtain the final result in (85).

$$0 \geq \inf_{\Theta_0^{\epsilon_n}} n^{-r} E[T_n(\theta)] \geq \inf_{\Theta_0^{\epsilon_n}} n^{1-r} h^{\frac{d_z}{2}} E\left[(E[m(X,\theta)|Z])^2 f_Z(Z) \right]$$

$$- \sup_{\Theta} n^{1-r} h^{\frac{d_z}{2}} \left| E[H_n(\theta)] - E\left[(E[m(X,\theta)|Z])^2 f_Z(Z) \right] \right| = O(n^{1-r} h^{\frac{d_z}{2} + \lfloor k \rfloor}) \tag{85}$$

Hence, since $nh^{\frac{d_z}{2} + \lfloor k \rfloor} \to 0$, the inequalities in (85) imply $\inf_{\Theta_0^{\epsilon_n}} E[T_n(\theta)] = o(1)$. Thus, applying Lemma .4 and inequality (83), we obtain (86).

$$\inf_{\Theta_0^{\epsilon_n} \cap R} T_n^*(\theta) + o_{as}(1) \leq \inf_{\Theta_0^{\epsilon_n} \cap R} \left(n^{-r} T_n(\theta) + T_n^*(\theta) \right) \leq \inf_{\Theta_0 \cap R} T_n^*(\theta) + o_{as}(1) \tag{86}$$

Next, notice that since Θ_0, R and $\Theta_0^{\epsilon_n}$ are closed and Θ is compact under $||\cdot||_{c\delta}$, as shown in Gallant & Nychka (1987), the infimums in the left hand side of (87) are attained. The first inequality in (87) is then implied for $\theta^* = \arg\min_{\Theta_0^{\epsilon_n}\cap R} T_n^*(\theta)$ and $\theta_p^* = \arg\min_{\Theta_0\cap R} ||\theta^* - \theta||_{c\delta}$. Next we argue $||\theta^* - \theta_p||_\infty \to 0$. For $\delta > 0$, define $A^\delta = \{\tilde{\theta} \in \Theta : \inf_{\Theta_0\cap R} ||\tilde{\theta} - \theta||_\infty \geq \delta\}$, which is compact under $||\cdot||_\infty$. Therefore, by continuity under $||\cdot||_\infty$, the minimum $\pi^* = \min_{A^\delta} E\left[(E[m(X,\theta)|Z])^2 f_Z(Z)\right]$ is attained with $\pi^* > 0$. Hence, since $A^\delta \cap \Theta_0^{\epsilon_n} \cap R = \emptyset$ for $\epsilon_n < \pi^*$ it follows that $\sup_{\theta_1\in\Theta_0^{\epsilon_n}\cap R} \inf_{\theta_2\in\Theta_0\cap R} ||\theta_1 - \theta_2||_\infty < \delta$ when $\epsilon_n < \pi^*$. Therefore, letting $\delta_n = ||\theta_p - \theta^*||_\infty$ implies the second inequality in (87).

$$\inf_{\Theta_0\cap R} T_n^*(\theta) - \inf_{\Theta_0^{\epsilon_n}\cap R} T_n^*(\theta) \leq T_n^*(\theta_p) - T_n^*(\theta^*)$$

$$\leq \sup_{||\theta_1-\theta_2||_\infty \leq \delta_n} |T_n^*(\theta_1) - T_n^*(\theta_2)| \tag{87}$$

Because the left hand side of (87) is always weakly positive, result (87) and the *a.s.* asymptotic equicontinuity of $T_n^*(\theta)$ under $||\cdot||_\infty$, established in Lemma 3.2, implies result (88).

$$\inf_{\Theta_0^{\epsilon_n}\cap R} T_n^*(\theta) \xrightarrow{p^*} \inf_{\Theta_0\cap R} T_n^*(\theta) \quad a.s. \tag{88}$$

Thus, combining the inequality from (86) and the convergence result in (88) establishes (89).

$$\inf_{\Theta_0^{\epsilon_n}\cap R} \left(n^{-r}T_n(\theta) + T_n^*(\theta)\right) \xrightarrow{p^*} \inf_{\Theta_0\cap R} T_n^*(\theta) \quad a.s. \tag{89}$$

To study the process on $(\Theta_0^{\epsilon_n})^c \cap R$, notice that the first inequality in (90) follows by direct calculation. Since $\sup_\Theta T_n^*(\theta) = O_{p^*}(1)$ *a.s.* by Lemma 3.2, Lemma .4 implies the final result in (90).

$$\inf_{(\Theta_0^{\epsilon_n})^c\cap R} \left(n^{-r}T_n(\theta) + T_n^*(\theta)\right) \geq \inf_{(\Theta_0^{\epsilon_n})^c\cap R} n^{-r}T_n(\theta) - \sup_\Theta T_n^*(\theta) \xrightarrow{p^*} \infty \quad a.s. \tag{90}$$

To conclude, notice that since $\inf_{\Theta_0\cap R} T_n^*(\theta) = O_{p^*}(1)$ *a.s.* by Lemma 3.2, results (89) and (90) establish (91).

$$\inf_{\Theta\cap R} \left(n^{-r}T_n(\theta) + T_n^*(\theta)\right) = \inf_{\Theta_0\cap R} T_n^*(\theta) + o_{p^*}(1) \quad a.s. \tag{91}$$

The continuous mapping theorem, result (91) and Lemma 3.2 then establish the first claim of the Theorem.

Now suppose $\Theta_0 \cap R = \emptyset$. Let $\pi^* = \inf_{\Theta\cap R} E[(E[m(X,\theta)|Z])^2 f_Z(Z)]$. By compactness of $\Theta\cap R$ and continuity under $||\cdot||_\infty$, the infimum is attained for some $\theta^* \notin \Theta_0$, which in turn implies $\pi^* > 0$. Therefore, $\Theta \cap R \subseteq (\Theta_0^{\epsilon_n})^c$ for ϵ_n small enough. Together with (90), this implies (92).

$$\inf_{\Theta\cap R} \left(n^{-r}T_n(\theta) + T_n^*(\theta)\right) \geq \inf_{(\Theta_0^{\epsilon_n})^c\cap R} \left(n^{-r}T_n(\theta) + T_n^*(\theta)\right) \xrightarrow{p^*} \infty \quad a.s. \tag{92}$$

which concludes the proof of the Theorem. ∎

Proof of Theorem 3.1: Throughout the proof I will use the same notations as in the proof of Theorem .3. We begin by establishing the first claim. First note that since Θ is compact under $||\cdot||_{c\delta}$, the different closedness assumptions imply $\Theta_j \cap R$ and $\Theta_0 \cap R$ are both compact under $||\cdot||_\infty$. Hence, θ^* and θ_p^* in (93) are well defined.

$$\theta^* = \arg\min_{\Theta_0\cap R} \left(n^{-r}T_n(\theta) + T_n^*(\theta)\right) \qquad \theta_p^* = \arg\min_{\Theta_j\cap R} ||\theta - \theta^*||_\infty \tag{93}$$

Furthermore, since $\theta^* \in \Theta_0$, $E[m(X,\theta^*)|z] = 0$, and therefore by Assumption 2(iii) inequality (94) follows.

$$E[(E[m(X,\theta_p^*)|Z])^2 f_Z(Z)] \leq E[G^2(X)f_Z(Z)]||\theta_p^* - \theta^*||_\infty^2 \tag{94}$$

24

Next, notice that $||\theta_p^* - \theta^*||_\infty = o((n^{1-r}h^{\frac{d_z}{2}})^{-1})$ and $n^{\frac{1}{2}-r}h^l \to \infty$ imply $||\theta_p^* - \theta^*||_\infty = o(h^{\frac{l}{2}})$. Hence, it follows by (94), that for n large enough $\theta_p^* \in \Theta_0^{\epsilon_n}$, for $\Theta_0^{\epsilon_n}$ as defined in (82). In turn, this implies the inequality in (95). The second result then holds for $\delta_n = \sup_{\Theta \cap R} \inf_{\Theta_j \cap R} ||\theta - \theta_j||_\infty$ due to (84) and Lemma .4.

$$n^{-r}T_n(\theta_p^*) \leq \sup_{\Theta_0^{\epsilon_n}} n^{-r}|T_n(\theta_p^*) - E[T_n(\theta_p^*)]| + n^{-r}E[T_n(\theta_p^*)] = o_{as}(1) + O(n^{1-r}h^{\frac{d_z}{2}}\delta_n^2 + n^{1-r}h^{\frac{d_z}{2}+\lfloor k \rfloor}) \quad (95)$$

The first inequality in (96) is then implied by (91). In order to obtain the second inequality then employ (95), $n^{1-r}h^{\frac{d_z}{2}}\delta_n^2 \to 0$ by Assumption 4(ii) and $n^{1-r}h^{\frac{d_z}{2}+\lfloor k \rfloor} \to 0$ by Assumption 4(i).

$$\inf_{\Theta \cap R}\left(n^{-r}T_n(\theta) + T_n^*(\theta)\right) - \inf_{\Theta_j \cap R}\left(n^{-r}T_n(\theta) + T_n^*(\theta)\right) \geq T_n^*(\theta^*) - \left(n^{-r}T_n(\theta_p^*) + T_n^*(\theta_p^*)\right) + o_{p^*}(1) \quad a.s.$$

$$\geq - \sup_{||\theta_1 - \theta_2||_\infty \leq \delta_n} |T_n^*(\theta_1) - T_n^*(\theta_2)| + o_{p^*}(1) \quad a.s. \quad (96)$$

Furthermore, since $\Theta_j \cap R \subseteq \Theta \cap R$ implies $\inf_{\Theta \cap R} n^{-r}T_n(\theta) + T_n^*(\theta) \leq \inf_{\Theta_j \cap R} n^{-r}T_n(\theta) + T_n^*(\theta)$, result (97) follows from (96) and Lemma 3.2.

$$\inf_{\Theta_j \cap R} n^{-r}T_n(\theta) + T_n^*(\theta) \xrightarrow{p^*} \inf_{\Theta \cap R} n^{-r}T_n(\theta) + T_n^*(\theta) \quad a.s. \quad (97)$$

Theorem .3 then implies the first claim of the theorem. For the second part, simply note that if $\Theta_0 \cap R = \emptyset$, then (98) follows by Theorem .3 and $\Theta_j \subseteq \Theta$.

$$\inf_{\Theta_j \cap R} n^{-r}T_n(\theta) + T_n^*(\theta) \geq \inf_{\Theta \cap R} n^{-r}T_n(\theta) + T_n^*(\theta) \xrightarrow{p^*} \infty \quad a.s. \quad (98)$$

Hence concluding the proof of the Theorem. ∎

Lemma .4. *Let* $\Theta_0^{\epsilon_n} = \left\{\theta \in \Theta : E\left[\left(E[m(X,\theta)|Z]\right)^2 f_Z(Z)\right] \leq \epsilon_n\right\}$, *where* $\epsilon_n = Ch^l$ *for* $l \leq \lfloor k \rfloor$, $n^{\frac{1}{2}-r}h^l \to \infty$ *and* $n^{\gamma+(1-2r)\frac{\bar u}{2}}h^{\left(d_z + l - \frac{l(m_0+\delta_0)d_x}{2m_0\delta_0}\right)\frac{\bar u}{2}} \to 0$ *for some* $\gamma > 1$. *If* C *is large enough, and Assumptions 1(i)-(iii), 2(i)-(iv), 3(i)-(ii) and 4(i) hold, then:*

$$\inf_{(\Theta_0^{\epsilon_n})^c \cap R} n^{-r}T_n(\theta) \xrightarrow{a.s.} \infty \qquad\qquad \sup_{\Theta_0^{\epsilon_n} \cap R} n^{-r}|T_n(\theta) - E[T_n(\theta)]| \xrightarrow{a.s.} 0$$

Proof of Lemma .4: To establish the first result, we define the U-Statistic:

$$C_n(\theta) \equiv \left(nh^{\frac{d_z}{2}}\right)^{-1} T_n(\theta) = \frac{2}{n(n-1)} \sum_{i=2}^{n} \sum_{j<i} H_n(w_i, w_j, \theta) \quad (99)$$

In addition, let $P_n^i(\theta)$ be the ith term in the Hoeffding decomposition of $C_n(\theta)$. The inequality in (100) then follows.

$$\inf_{(\Theta_0^{\epsilon_n})^c \cap R} n^{-r}T_n(\theta) \geq (nh^{d_z})^{\frac{1}{2}} \left(n^{\frac{1}{2}-r} \inf_{(\Theta_0^{\epsilon_n})^c} E[H_n(\theta)] - \sup_\Theta n^{\frac{1}{2}-r}|P_n^1(\theta)|\right) - \sup_\Theta n^{1-r}h^{\frac{d_z}{2}}|P_n^2(\theta)| \quad (100)$$

Next, define the class $\tilde\Theta_n^1 = \{E[H_n(\theta)|w_i] - E[H_n(\theta)] : \theta \in \Theta\}$ and notice that Theorem .1 implies the first inequality in (101) for D_n the diameter of $\tilde\Theta_n^1$ under $\rho_{1n}(\theta_1, \theta_2)$. Furthermore, by Lemma .3, $|\tilde\theta_1(w_i) - \tilde\theta_2(w_i)| \leq G_{1n}(w_i)||\theta_1 - \theta_2||_\infty$ for any $\tilde\theta_1, \tilde\theta_2 \in \tilde\Theta_n^1$ and $G_{1n}(w_i) = E[J_n|w_i] + E[J_n]$. The second inequality in (101) then follows by Theorem .2 and the change of variables $u = \epsilon/2||G_{1n}||_{1n}$. Theorem 0.4 in Santos (2007) and $D_n \leq M||G_{1n}||_{1n}$ for some M not depending on n implies the third result in (101). Use convexity due to $\bar u \geq 2$ to obtain the final result

in (101).

$$E\left[\left(\sup_{\Theta} n^{\frac{1}{2}-r}|P_n^1(\theta)|\right)^{\bar{u}}\right] \lesssim n^{-r\bar{u}} E\left[\left(\int_0^{D_n} [\log N(\tilde{\Theta}_n^1, \rho_{1n}, \epsilon)]^{\frac{1}{2}} d\epsilon\right)^{\bar{u}}\right]$$

$$\lesssim n^{-r\bar{u}} E\left[\|G_{1n}\|_{1n}^{\bar{u}} \left(\int_0^{\frac{D_n}{\|G_{1n}\|_{1n}}} [\log N(\Theta, \|\cdot\|_{\infty}, u)]^{\frac{1}{2}} du\right)^{\bar{u}}\right]$$

$$\lesssim n^{-r\bar{u}} E\left[\left(\frac{1}{n}\sum_{i=1}^n (E[J_n|W_i] + E[J_n])^2\right)^{\frac{\bar{u}}{2}}\right]$$

$$\leq n^{-r\bar{u}} E\left[(E[J_n|W_i] + E[J_n])^{\bar{u}}\right] \quad (101)$$

Therefore, using Markov's inequality we obtain the first inequality in (102). Next, note that Assumption 4(i) and Lemma .3 imply $(E[J_n])^{\bar{u}}$ and $E[(E[J_n|W_i])^{\bar{u}}]$ are uniformly bounded in n. Since $r\bar{u} > 1$, the final inequality follows.

$$\sum_{i=1}^{\infty} P\left(\sup_{\Theta} n^{\frac{1}{2}-r}|P_n^1(\theta)| > \epsilon\right) \leq \frac{1}{\epsilon^{\bar{u}}}\sum_{i=1}^{\infty} \frac{1}{n^{r\bar{u}}} E\left[(E[J_n|W_i] + E[J_n])^{\bar{u}}\right] < \infty \quad (102)$$

In turn, (102) and the Borel-Cantelli Lemma imply (103).

$$n^{\frac{1}{2}-r}\sup_{\Theta}|P_n^1(\theta)| \xrightarrow{a.s.} 0 \quad (103)$$

Next, note that the first inequality in (104) follows by the same arguments as in (102). The second inequality is then implied by convexity, since $\bar{u} \geq 2$.

$$E\left[\left(\sup_{\Theta} n^{1-r}h^{\frac{d_z}{2}}|P_n^2(\theta)|\right)^{2\lfloor\frac{\bar{u}}{2}\rfloor}\right]$$

$$\lesssim n^{-r2\lfloor\frac{\bar{u}}{2}\rfloor}h^{d_z\lfloor\frac{\bar{u}}{2}\rfloor} E\left[\left(\frac{2}{n(n-1)}\sum_{i=2}^n\sum_{j<i}(J_n(W_i,W_j) + E[J_n|W_i] + E[J_n|W_j] + E[J_n])^2\right)^{\lfloor\frac{\bar{u}}{2}\rfloor}\right]$$

$$\lesssim n^{-r2\lfloor\frac{\bar{u}}{2}\rfloor}h^{d_z\lfloor\frac{\bar{u}}{2}\rfloor} E\left[\left(\frac{2}{n(n-1)}\sum_{i=2}^n\sum_{j<i}J_n^2(W_i,W_j)\right)^{\lfloor\frac{\bar{u}}{2}\rfloor}\right] + n^{-r2\lfloor\frac{\bar{u}}{2}\rfloor}h^{d_z\lfloor\frac{\bar{u}}{2}\rfloor} E\left[(E[J_n|W_i] + E[J_n])^{2\lfloor\frac{\bar{u}}{2}\rfloor}\right]$$

$$(104)$$

Let A be the set of vectors $\lambda = (\lambda_1, \ldots, \lambda_{n(n-1)/2})$ such that each λ_k is a nonnegative integer and $\sum_k \lambda_k = \lfloor\frac{\bar{u}}{2}\rfloor$. The first equality in (105) then follows from the multinomial theorem. Next, let $A(l,s)$ be the set of $\lambda \in A$ such that l indices have $\lambda_k \neq 0$, and these l pairs are formed by s distinct W_i. Let $c(l)$ be the smallest integer m satisfying $m(m-1) \geq 2l$, and note that this is the smallest number of distinct W_i that can form l distinct pairs. The second inequality in (105) then follows by noting $A = \cup_{s,l} A(s,l)$.

$$E\left[\left(\frac{2}{n(n-1)}\sum_{i=2}^n\sum_{j<i}J_n^2(W_i,W_j)\right)^{\lfloor\frac{\bar{u}}{2}\rfloor}\right]$$

$$= \left(\frac{2}{n(n-1)}\right)^{\lfloor\frac{\bar{u}}{2}\rfloor}\sum_{\lambda\in A}\frac{\lfloor\frac{\bar{u}}{2}\rfloor!}{\lambda_1!\ldots\lambda_{n(n-1)/2}!}E\left[J^{2\lambda_1}(W_2,W_1)\ldots J^{2\lambda_{n(n-1)/2}}(W_n,W_{n-1})\right]$$

$$\lesssim \left(\frac{2}{n(n-1)}\right)^{\lfloor\frac{\bar{u}}{2}\rfloor}\sum_{l=1}^{\lfloor\frac{\bar{u}}{2}\rfloor}\sum_{s=c(l)}^{2l}\sum_{\lambda\in A(l,s)}E\left[J^{2\lambda_1}(W_2,W_1)\ldots J^{2\lambda_{n(n-1)/2}}(W_n,W_{n-1})\right] \quad (105)$$

26

In turn, use the definition of $J_n(w_i, w_j)$ in (79), the $K(u)$, $E[G^j(X)|z]$ and $E[F^j(X)|z]$ being bounded for $1 \le j \le \bar{u}$, to obtain the inequality in (106). Define the resulting expression to be $e(\lambda)$.

$$E\left[J^{2\lambda_1}(W_2, W_1) \ldots J^{2\lambda_{n(n-1)/2}}(W_n, W_{n-1})\right] \lesssim h^{-2d_z\lfloor \frac{\bar{u}}{2} \rfloor} E\left[K^{2\lambda_1}\left(\frac{Z_1 - Z_2}{h}\right) \ldots K^{2\lambda_{n(n-1)/2}}\left(\frac{Z_n - Z_{n-1}}{h}\right)\right]$$

$$\equiv h^{-2d_z\lfloor \frac{\bar{u}}{2} \rfloor} e(\lambda) \tag{106}$$

For each λ let $b(\lambda)$ be the largest integer k such that $h^{-d_z k} e(\lambda) = O(1)$. Note that the value of $b(\lambda)$ will depend on how many substitutions of the form $u = (z_i - z_j)/h$ are appropriate when calculating $e(\lambda)$. Similarly, we also define:

$$B(l, s) = \min_{\lambda \in A(l,s)} b(\lambda) \tag{107}$$

Since the cardinality of $A(l, s)$ is smaller than $\binom{n}{s}$, using (105), (106) and (107) then implies the first inequality in (108). Notice, however, that $B(l, s+1) \le B(l, s) + 1$ and similarly $B(l+1, 2l+2) \le B(l, 2l) + 1$. Thus, since $nh^{d_z} \to \infty$, the term of largest order in the summation corresponds to $(l, s) = (\lfloor \frac{\bar{u}}{2} \rfloor, 2\lfloor \frac{\bar{u}}{2} \rfloor)$. Using $B(\lfloor \frac{\bar{u}}{2} \rfloor, 2\lfloor \frac{\bar{u}}{2} \rfloor) = \lfloor \frac{\bar{u}}{2} \rfloor$ then establishes the second inequality in (108).

$$E\left[\left(\frac{2}{n(n-1)} \sum_{i=2}^{n} \sum_{j<i} J_n^2(W_i, W_j)\right)^{\lfloor \frac{\bar{u}}{2} \rfloor}\right] \lesssim \left(\frac{2}{n(n-1)}\right)^{\lfloor \frac{\bar{u}}{2} \rfloor} \sum_{l=1}^{\lfloor \frac{\bar{u}}{2} \rfloor} \sum_{s=c(l)}^{2l} O\left(\binom{n}{s} \times h^{-d_z(2\lfloor \frac{\bar{u}}{2} \rfloor - B(l,s))}\right)$$

$$\lesssim O\left(n^{-\lfloor \frac{\bar{u}}{2} \rfloor}\right) \times O\left(n^{\lfloor \frac{\bar{u}}{2} \rfloor} \times h^{-d_z(2\lfloor \frac{\bar{u}}{2} \rfloor - \lfloor \frac{\bar{u}}{2} \rfloor)}\right) \tag{108}$$

Hence, since by Lemma .3, $(E[J_n])^{2\lfloor \frac{\bar{u}}{2} \rfloor}$ and $E[(E[J_n|W_i])^{2\lfloor \frac{\bar{u}}{2} \rfloor}]$ are uniformly bounded in probability, combining (104) and (108) establishes (109).

$$E\left[\left(\sup_{\Theta} n^{1-r} h^{\frac{d_z}{2}} |P_n^2(\theta)|\right)^{2\lfloor \frac{\bar{u}}{2} \rfloor}\right] = O\left(n^{-2r\lfloor \frac{\bar{u}}{2} \rfloor}\right) \tag{109}$$

Arguing as in (102) and (103) together with $2r\lfloor \frac{\bar{u}}{2} \rfloor > 1$ in turn establishes (110).

$$n^{1-r} h^{\frac{d_z}{2}} \sup_{\Theta} |P_n^2(\theta)| \xrightarrow{a.s.} 0 \tag{110}$$

Furthermore, it follows that the first inequality in (111) holds for M sufficiently large due to result (84). Next, notice that $\epsilon_n - M h^{\lfloor k \rfloor} \ge \epsilon_n / 2$ by setting the level of ϵ_n sufficiently large when $\epsilon_n = h^{\lfloor k \rfloor}$, or trivially when $\epsilon_n = h^l$ with $l < \lfloor k \rfloor$. The final result in (111) then follows by $n^{\frac{1}{2}-r} \epsilon_n \to \infty$.

$$n^{\frac{1}{2}-r} \inf_{(\Theta_0^{\epsilon n})^c \cap R} E[H_n(\theta)] \ge n^{\frac{1}{2}-r}(\epsilon_n - M h^{\lfloor k \rfloor}) \to \infty \tag{111}$$

Combining (100), (103), (110) and (111) then establishes the first claim of the Lemma.

In order to establish the second claim of the Lemma, we derive a useful inequality in (112) which immediately follows from the Hoeffding decomposition of $(T_n(\theta) - E[T_n(\theta)])$.

$$\sup_{\Theta_0^{\epsilon n} \cap R} n^{-r} |T_n(\theta) - E[T_n(\theta)]| \le \sup_{\Theta_0^{\epsilon n}} n^{1-r} h^{\frac{d_z}{2}} |P_n^1(\theta)| + \sup_{\Theta} n^{1-r} h^{\frac{d_z}{2}} |P_n^2(\theta)| \tag{112}$$

Next, notice that the first inequality in (113) was already established in (101). For the second inequality use Theorem 0.4 in Santos (2007), $G_{1n}(w_i)$ as defined in the derivation in (101) and let $\lambda = (m_0 + \delta_0) d_x/(2m_0 \delta_0)$. Holder's inequality in turn implies the third inequality in (113).

$$E\left[\left(\sup_{\Theta_0^{\epsilon n}} n^{1-r} h^{\frac{d_z}{2}} |P_n^1(\theta)|\right)^{\bar{u}}\right] \lesssim n^{(\frac{1}{2}-r)\bar{u}} h^{\frac{d_z}{2}\bar{u}} E\left[||G_{1n}||_{1n}^{\bar{u}} \left(\int_0^{\frac{D_n}{||G_{1n}||_{1n}}} [\log N(\Theta, || \cdot ||_\infty, u)]^{\frac{1}{2}} du\right)^{\bar{u}}\right]$$

$$\lesssim n^{(\frac{1}{2}-r)\bar{u}} h^{\frac{d_z}{2}\bar{u}} E\left[D_n^{\bar{u}(1-\lambda)} ||G_{1n}||_{1n}^{\bar{u}\lambda}\right]$$

$$\le n^{(\frac{1}{2}-r)\bar{u}} h^{\frac{d_z}{2}\bar{u}} \left(E\left[D_n^{\bar{u}}\right]^{1-\lambda}\right) \left(E\left[||G_{1n}||_{1n}^{\bar{u}}\right]\right)^\lambda \tag{113}$$

To examine the right hand side of (113), in (114) we first study $E[D_n^{\bar{u}}]$. The first equality in (114) follows from the definition of D_n, while the first and second inequalities follows by convexity as $\bar{u} \geq 2$.

$$E[D_n^{\bar{u}}] = E\left[\sup_{\theta_1,\theta_2 \in \Theta_0^{\epsilon_n}} \left(\frac{1}{n}\sum_{i=1}^n (E[H_n(\theta_1)|w_i] - E[H_n(\theta_1)] - (E[H_n(\theta_2)|w_i] - E[H_n(\theta_2)])^2\right)^{\frac{\bar{u}}{2}}\right]$$

$$\leq 2^{\frac{\bar{u}}{2}} E\left[\sup_{\Theta_0^{\epsilon_n}} \left(\frac{1}{n}\sum_{i=1}^n (E[H_n(\theta)|w_i] - E[H_n(\theta)])^2\right)^{\frac{\bar{u}}{2}}\right]$$

$$\leq 2^{\bar{u}} E\left[\sup_{\Theta} \left|\frac{1}{n}\sum_{i=1}^n (E[H_n(\theta)|w_i] - E[H_n(\theta)])^2 - \text{Var}(E[H_n(\theta)|W_i])\right|^{\frac{\bar{u}}{2}}\right] + 2^{\bar{u}}\sup_{\Theta_0^{\epsilon_n}} \text{Var}(E[H_n(\theta)|W_i])^{\frac{\bar{u}}{2}} \quad (114)$$

Next, define the function $G_{3n}(w_i) = 2(E[J_n|w_i])^2 + 4E[(E[J_n|W_i])^2]$. Using Lemma .1 and the arguments in (101) then implies the inequality in (115).

$$E\left[\sup_{\Theta} \left|\frac{1}{\sqrt{n}}\sum_{i=1}^n (E[H_n(\theta)|w_i] - E[H_n(\theta)])^2 - \text{Var}(E[H_n(\theta)|W_i])\right|^{\frac{\bar{u}}{2}}\right] \lesssim E[\|G_{3n}\|_{1n}^{\frac{\bar{u}}{2}}] \quad (115)$$

To control the right hand side of (114) we now examine $E(E[H_n(\theta)|w_i])^2)$ in (116). The first equality follows by inspection, while the first inequality is implied by Assumption 2(ii) and the change of variables $u = (z_i - z_j)/h$. The final equality then follows by a Taylor expansion of order $\lfloor k \rfloor$ at $h = 0$ and Assumptions 2(i) and 3(i).

$$E[(E[H_n(\theta)|w_i])^2] = h^{-2d_z} E\left[E[m^2(X_i,\theta)|Z_i]\left(E\left[K\left(\frac{Z_i - Z_j}{h}\right) E[m(X_j,\theta)|Z_j]\Big|Z_i\right]\right)^2\right]$$

$$\lesssim \int \left(\int K(u) E[m(X_i,\theta)|z_i - hu] f_Z(z_i - hu) du\right)^2 f_Z(z_i) dz_i$$

$$= E[(E[m(X,\theta)|Z])^2 f_Z^2(Z)] + O(h^{\lfloor k \rfloor}) \quad (116)$$

By Lemma .3, $E[\|G_{3n}\|_{1n}^{\frac{\bar{u}}{2}}]$ is uniformly bounded in n, while (116) implies $\sup_{\Theta_0^{\epsilon_n}} \text{Var}(E[H_n(\theta)|W_i]) = O(\epsilon_n + h^{\lfloor k \rfloor})$. Hence, combining (114), (115) and $\epsilon_n = h^l$ with $\sqrt{n}h^l \to \infty$ and $l \leq \lfloor k \rfloor$ implies (117).

$$E[D_n^{\bar{u}}] = O(h^{l\frac{\bar{u}}{2}}) \quad (117)$$

Furthermore, notice that Lemma .3 implies $E[\|G_{1n}\|_{1n}^{\bar{u}}]$ is uniformly bounded in n. Therefore, combining results (113) and (117) we derive the equality in (118).

$$E\left[\left(\sup_{\Theta_0^{\epsilon_n}} n^{1-r} h^{\frac{d_z}{2}} |P_n^1(\theta)|\right)^{\bar{u}}\right] = O\left(n^{(\frac{1}{2}-r)\bar{u}} h^{(d_z + l(1-\lambda))\frac{\bar{u}}{2}}\right) \quad (118)$$

By Assumption 4(i) there exists a $\gamma > 1$ such that $n^{\gamma + (1-2r)\frac{\bar{u}}{2}} h^{\left(d_z + l - \frac{l(m_0 + \delta_0)d_x}{2m_0\delta_0}\right)\frac{\bar{u}}{2}} \to 0$. Therefore, arguing as in (102) and (103) establishes (119).

$$\sup_{\Theta_0^{\epsilon_n}} n^{1-r} h^{\frac{d_z}{2}} |P_n^1(\theta)| \xrightarrow{a.s.} 0 \quad (119)$$

Combining results (112), (119) and (110) in turn establishes the second claim of the Lemma. ∎

Proof of Theorem 3.2: In order to show the first claim we begin by establishing that the law of

$$\min_{\Theta_0 \cap R} G(\theta) \quad (120)$$

is absolutely continuous with respect to Lebesgue measure. Define, $f(\xi) : \mathcal{L}^\infty(\Theta_0) \to \Re$ by $f(\xi) = -\min_{\Theta_0 \cap R} \xi(\theta)$ for $\xi(\theta) \in \mathcal{L}^\infty(\Theta_0)$. Notice that $f(\xi)$ is a convex functional, and in addition it is continuous with respect to $\|\cdot\|_\infty$.

Furthermore, by tightness, $P(f(G(\theta)) = -\infty) = 0$. This verifies the conditions for Theorem 11.1 in Davydov, Lifshits & Smorodina (1998), which implies the law of $-\min_{\Theta_0 \cap R} G(\theta)$ is absolutely continuous with respect to Lebesgue measure. Thus, so is the law of $\min_{\Theta_0 \cap R} G(\theta)$. To conclude, note that since

$$I_n^*(R) \xrightarrow{\mathcal{L}^*} \min_{\Theta_0 \cap R} G(\theta) \quad a.s. \tag{121}$$

by Theorem .3, continuity of the limiting distribution and Theorem 1 in Beran (1984) establish the first claim.

For the second claim notice that the first equality in (122) follows by the definitions of $I_n(R)$ and $I_n^*(R)$, while the first and second inequalities are implied by simple manipulations and $\Theta_j \subseteq \Theta$.

$$
\begin{aligned}
P^*(I_n^*(R) \leq I_n(R)) &= P^* \left(\inf_{\Theta_j \cap R} \left(n^{-r} T_n(\theta) + T_n^*(\theta) \right) \leq \inf_{\Theta_j \cap R} T_n(\theta) \right) \\
&\geq P^* \left(\inf_{\Theta_j \cap R} n^{-r} T_n(\theta) + \sup_{\Theta_j \cap R} T_n^*(\theta) \leq \inf_{\Theta_j \cap R} T_n(\theta) \right) \\
&\geq P^* \left(\sup_{\Theta} T_n^*(\theta) \leq (1 - n^{-r}) \inf_{\Theta \cap R} T_n(\theta) \right)
\end{aligned}
\tag{122}
$$

Let $\pi^* = \inf_{\Theta \cap R} E[(E[m(X, \theta)|Z])^2 f_Z(Z)]$ and notice that by compactness of $\Theta \cap R$ the infimum is attained. In turn, $\Theta_0 \cap R = \emptyset$ implies that $\pi^* > 0$. Thus, $\Theta \cap R \subseteq (\Theta_0^{\epsilon_n})^c$ for ϵ_n sufficiently small and $\Theta_0^{\epsilon_n}$ as defined in (82). The inequality in (123) then follows for n sufficiently large since $r > 0$. The final result in (123) is implied by Lemma .4.

$$(1 - n^{-r}) \inf_{\Theta \cap R} T_n(\theta) \geq n^{-r} \inf_{(\Theta_0^{\epsilon_n})^c} T_n(\theta) \xrightarrow{as} \infty \tag{123}$$

By Lemma 3.2, $\sup_{\Theta} T_n^*(\theta) = O_{p^*}(1)$ almost surely, and therefore combining (122) and (123) we conclude (124).

$$P^*(I_n^*(R) \leq I_n(R)) \xrightarrow{as} 1 \tag{124}$$

Result (124) and the definition of $\hat{c}_{1-\alpha}$ in turn establish the second claim of the Theorem. ■

References

[1] AI, C. AND CHEN, X., "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions". *Econometrica*, 71, pp. 1795-1844, (2003).

[2] ARCONES, M. A. AND GINE, E., "On the Bootstrap of U and V Statistics". *Annals of Statistics*, 20, pp. 655-674, (1992).

[3] ARCONES, M. A. AND GINE, E., "Limit Theorems for U-Processes". *Annals of Probability*, 21, pp. 1494-1542, (1993).

[4] ARCONES, M. A. AND GINE, E., "U-Processes indexed by Vapnik-Červonenkis classes of functions with Applications to Asymptotics and Bootstrap of U-Statistics with Estimated Parameters". *Stochastic Processes and Their Applications*, 52, pp. 17-38, (1994).

[5] BERAN, R., "Bootstrap Methods in Statistics". *Jahresberichte des Deutschen Mathematischen Vereins*, 86, pp. 14-30, (1984).

[6] BLUNDELL, R., CHEN, X. AND KRISTENSEN, D., "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves". Working paper, Department of Economics, University College London, (2004).

[7] BORELL, C., "On the Integrability of Banach Space Valued Walsh Polynomials". *Seminaire de Probabilities*, 13, pp. 1-3, (1979).

[8] BUGNI, F. A., "Bootstrap Inference fin Partially Identified Models". Working Paper, Department of Economics, Northwestern University, (2008).

[9] CANAY, I. A., "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity". Working Paper, Department of Economics, Northwestern University, (2008).

[10] CHEN, X., "Large Sample Sieve Estimation of Semi-Nonparametric Models". Working paper, Department of Economics, New York Unversity (2006).

[11] CHEN, X. AND SHEN, X., "Sieve Extremum Estimates for Weakly Dependent Data". *Econometrica*, 66, pp. 289-314, (1998).

[12] DAROLLES, S., FLORENS, J. AND RENAULT, E., "Nonparametric Instrumental Regression". Working paper, GREMAQ, University of Tolouse, (2003).

[13] DAVYDOW, Y. A., LIFSHITS, M. A. AND SMORODINA, N.V., "Local Properties of Distributions of Stochastic Functionals". American Mathematical Society, (1998).

[14] DE LA PENA, V. H., "Decoupling and Khintchine's Inequalities for U-Statistics". *Annals of Probability*, 20, pp. 1877-1892, (1992).

[15] GAGLIARDINI, P. AND SCAILLET, O., "A Specification Test for Nonparametric Instrumental Variable Regression". Working Paper, Department of Economics, University of Lugano, (2007a).

[16] GAGLIARDINI, P. AND SCAILLET, O., "Tikhonov Regularization for Nonparametric Instrumental Variable Estimators". Working Paper, Department of Economics, University of Lugano, (2007b).

[17] GALLANT, A. R. AND NYCHKA, D. W., "Semi-Nonparametric Maximum Likelihood Estimation". *Econometrica*, 55, pp.363-390, (1987).

[18] HALL, P., "Integrated Square Error Properties of Kernel Estimators of Regression Functions". *Annals of Statistics*, 12, pp. 241-260, (1984).

[19] HALL, P. AND HOROWITZ, J., "Nonparametric Methods for Inference in the Presense of Instrumental Variables". *Annals of Statistics*, 33, pp. 2904-2929, (2005).

[20] HOROWITZ, J., "Testing a Parametric Model Against a Nonparametric Alternative with Identification through Instrumental Variables". *Econometrica*, 71, pp. 521-538, (2006).

[21] HOROWTIZ, J., "Asymptotic Normality of a Nonparametric Instrumental Variables Estimator". Working Paper, Department of Economics, Northwestern University, (2007).

[22] IMBENS, G. W. AND MANSKI, C. F., "Confidence Intervals for Partially Identified Parameters". *Econometrica*, 72, pp. 1845-1857, (2004).

[23] IMBENS, G. W. AND NEWEY W. K., "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity". Working paper, Department of Economics, Harvard, (2006).

[24] LI, Q. AND WANG, S., "A Simple Consistent Bootstrap Test for a Parametric Regression Function". *Journal of Econometrics*, 87, pp. 145-165, (1998).

[25] MANSKI, C. F., "Nonparametric Bounds of Treatment Effects". *American Economic Review, Paper and Proceedings*, 80, pp. 319-323 (1990).

[26] MANSKI, C. F., "Partial Identification of Probability Distributions". Springer, New York (2003).

[27] NEWEY, W. AND POWELL, J., "Instrumental Variable Estimation of Nonparametric Models". *Econometrica*, 71, pp. 1565-1578, (2003).

[28] NEWEY, W. K., POWELL, J. L. AND VELLA F., "Nonparametric Estimation of Triangular Simultaneous Equation Models". *Econometrica*, 67, pp. 565-603, (1999).

[29] POLLARD, D., "Convergence of Stochastic Processes". Springer-Verlag, New York, (1984).

[30] ROMANO, J. P. AND SHAIKH, A. M., "Inference for Identifiable Parameters in Partially Identified Eonometric Models". Forthcoming, *Journal of Statistical Planning and Inference*.

[31] SANTOS, A., "Inference in Nonparametric Instrumental Variables with Partial Identification". Working paper, Department of Economics, University of California at San Diego, (2007).

[32] SANTOS, A., "Instrumental Variable Methods for Recovering Continuous Linear Functionals". Working paper, Department of Economics, University of California at San Diego (2008).

[33] SCHENNACH, S., CHALAK, K. AND WHITE, H. "Estimating Average Marginal Effects in Nonseparable Structural Systems". Working paper, Department of Economics, University of California at San Diego, (2007).

[34] SEVERINI, T. A. AND TRIPATHI, G., "Some Identification Issues in Nonparametric Linear Models with Endogenous Regressors". *Econometric Theory*, 22, pp. 258-278, (2006).

[35] SEVERINI, T. A. AND TRIPATHI, G., "Efficiency Bounds for Estimating Linear Functionals of Nonparametric Regression Models with Endogenous Regressors". Working Paper, University of Connecticut, (2007).

[36] VAN DER VAART, A. W. AND WELLNER, J., "Weak Convergence and Empirical Processes - With Applications to Statistics". Springer, New York, (1996).

[37] ZHENG, J. X., "A Consistent Test of Functional Form via Nonparametric Estimation Techniques". *Journal of Econometrics*, 75, pp. 263-289, (1996).